

# EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

---

Volume VII

1947

917 FIFTEENTH ST., N. W., • WASHINGTON 5, D. C.

# EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

FREDERIC KUDER, Editor

## ASSOCIATE EDITORS

DOROTHY C. ADKINS ..... United States Civil Service Commission  
FORREST A. KINGSBURY ..... University of Chicago  
FRED MCKINNEY, Editorial Representative of the American College  
Personnel Association ..... University of Missouri  
M. W. RICHARDSON ..... Richardson, Bellows, Henry and Co.

## BOARD OF COOPERATING EDITORS

JOHN G. DARLEY <i>University of Minnesota</i>	DAVID SEGEL <i>U. S. Office of Education</i>
HAROLD A. EDGERTON <i>Richardson, Bellows, Henry and Co.</i>	C. L. SHARTLE <i>Ohio State University</i>
MAX D. ENGELHART <i>Chicago City Junior Colleges</i>	H. C. TAYLOR <i>The W. E. Upjohn Institute for Community Research</i>
E. B. GREENE <i>United States Employment Service</i>	THELMA G. THURSTONE <i>University of Chicago</i>
J. P. GUILFORD <i>University of Southern California</i>	HERBERT A. TOOPS <i>Ohio State University</i>
E. F. LINDQUIST <i>State University of Iowa</i>	E. G. WILLIAMSON <i>University of Minnesota</i>
CHARLES I. MOSIER <i>Personnel Research Section, A.G.O.</i>	BEN D. WOOD <i>Columbia University</i>
P. J. RULON <i>Harvard University</i>	JOHN R. YALE <i>Science Research Associates</i>

This journal is open to (1) discussions of problems in the field of the measurement of individual differences, (2) reports of research on the development and use of tests and measurements in education, industry, and government, (3) descriptions of testing programs being used for various purposes, and (4) miscellaneous notes pertinent to the measurement field, such as suggestions of new types of items or improved methods of treating test data. Contributors receive one hundred reprints of their articles without charge. Manuscripts should be sent to Frederic Kuder, 917 Fifteenth Street, N. W., Washington 5, D. C. Writers are requested to include a biographical sketch with each manuscript, following the style of the section on contributors published in each issue.

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT is published quarterly, one volume per calendar year, at N. Queen St. and McGovern Ave., Lancaster, Pa. and 917 Fifteenth St., N. W. Washington 5, D. C. Entered as second class matter October 8, 1945 at the Post Office at Lancaster, Pa., under the act of March 3, 1879. Copyright, 1947, by Frederic Kuder.

Subscription rate \$4.00 a year, domestic and foreign. Single copies, \$1.25 with the exception of Volume VII, No. 3 (in two parts), for which the price is \$2.25. Back volumes: Volumes V (1945), VI (1946) and VII (1947), \$5.00 each. Volumes I through IV are available in a small-print edition at \$3.00 per volume (paper bound).



# INDEX FOR VOLUME VII

<i>Allen, Lucille</i>	
COORDINATION OF STUDENT PERSONNEL SERVICES AT CORNELL UNIVERSITY .....	655
<i>Altus, William D. (with Hugh M. Bell)</i>	
AN ANALYSIS OF FOUR ORALLY ADMINISTERED MEASURES OF ADJUSTMENT .....	101
<i>Andrews, Avonne L. (with Robert P. Fischer)</i>	
A STUDY OF THE EFFECT OF CONFORMITY TO SOCIAL EXPECTANCY ON EVALUATIVE ATTITUDES .....	331
<i>Beery, George S.</i>	
RELATIONSHIP OF THE COLLEGE ADMINISTRATION WITH FRATERNITIES AND SORORITIES .....	594
<i>Bell, Hugh M. (with William D. Altus)</i>	
AN ANALYSIS OF FOUR ORALLY ADMINISTERED MEASURES OF ADJUSTMENT .....	101
<i>Bergstresser, John</i>	
THE NATIONAL STUDENT MOVEMENT AND CONFERENCE ....	662
<i>Blyley, Katherine</i>	
THE KEUKA COLLEGE FIELD WORK PLAN .....	665
<i>Borreson, B. James</i>	
THE APPLICATION OF PERSONNEL METHODS TO UNIVERSITY HOUSING PROCEDURES .....	583
<i>Britt, Steuart Henderson (with Harold A. Edgerton)</i>	
TECHNICAL ASPECTS OF THE FOURTH ANNUAL SCIENCE TALENT SEARCH .....	3
<i>Brody, William (with Norman J. Powell)</i>	
A NEW APPROACH TO ORAL TESTING .....	289
<i>Callis, Robert (with C. Gilbert Wrenn)</i>	
THE GED TESTS AS PREDICTORS OF SCHOLASTIC SUCCESS ..	93
<i>Cattell, Raymond B.</i>	
THE ERGIC THEORY OF ATTITUDE AND SENTIMENT MEASUREMENT .....	221
<i>Collins, Duane</i>	
AN IN-SERVICE TRAINING PROGRAM FOR RESIDENCE COUNSELORS .....	647
<i>Cook, Walter W. (with Carroll A. Leeds)</i>	
MEASURING THE TEACHING PERSONALITY .....	399
<i>DuBois, Philip H. (with Vernon W. Lemmon)</i>	
MULTIPLE FORMS OF OBJECTIVE TESTS: A TEST-SCORING MACHINE TECHNIQUE .....	783
<i>Durost, Walter N.</i>	
WHAT CONSTITUTES A MINIMAL SCHOOL TESTING PROGRAM	45

<i>Ebel, Robert L.</i>	
THE FREQUENCY OF ERRORS IN THE CLASSIFICATION OF INDIVIDUALS ON THE BASIS OF FALLIBLE TEST SCORES .....	725
<i>Edgerton, Harold A. (with Stuart Henderson Brill)</i>	
TECHNICAL ASPECTS OF THE FOURTH ANNUAL SCIENCE TALENT SEARCH .....	3
<i>Ekberg, David L.</i>	
A STUDY IN TOOL USAGE .....	421
<i>Englehart, Max D.</i>	
SUGGESTIONS FOR WRITING ACHIEVEMENT EXERCISES TO BE USED IN TESTS SCORED ON THE ELECTRIC SCORING MACHINE .....	357
<i>Evans, M. Catharine</i>	
SOCIAL ADJUSTMENT AND INTEREST SCORES OF INTROVERTS AND EXTROVERTS .....	157
<i>Feder, Daniel D.</i>	
ESTIMATION OF TEST ITEM DIFFICULTY BY AVERAGING HIGHEST AND LOWEST QUARTER PERFORMANCE COMPARED WITH TOTAL POPULATION COUNT .....	133
<i>Feder, Daniel D.</i>	
PRESIDENTIAL ADDRESS: WHEN COLLEGES BULGE .....	475
<i>Fischer, Robert P. (with Avonne L. Andrews)</i>	
A STUDY OF THE EFFECT OF CONFORMITY TO SOCIAL EXPECTANCY ON EVALUATIVE ATTITUDES .....	331
<i>Foley, John D.</i>	
DISCIPLINE: A STUDENT COUNSELING APPROACH .....	569
<i>Frœhlich, Clifford P. (with Robert J. Keller)</i>	
A TECHNIQUE FOR OBTAINING SUMMARY DATA FROM AGGREGATE WEIGHTING SHEETS ON THE GRAPHIC ITEM COUNTER ATTACHMENT TO THE TEST SCORING MACHINE ..	429
<i>Garrett, Wiley S.</i>	
THE NUMBER RELATIONS SECTION OF THE CARNEGIE MENTAL ABILITY TESTS TREATED AS A POWER TEST .....	309
<i>Gaylord, R. H. (with E. K. Taylor)</i>	
TABLE FOR USE IN THE COMPUTATION OF STATISTICS OF DICHOTOMOUS AND TRUNCATED DISTRIBUTIONS .....	441
<i>Gregory, Wilbur S.</i>	
DEFECTS IN EDUCATIONAL AND PERSONNEL SYSTEMS AS REFLECTED IN MILITARY EXPERIENCES .....	505
<i>Guttman, Louis</i>	
THE CORNELL TECHNIQUE FOR SCALE AND INTENSITY ANALYSIS .....	247
<i>Harmon, Lindsey R.</i>	
TEST PATTERNS IN THE VOCATIONAL CLINIC .....	207
<i>Harris, C. W.</i>	
PREDICTION OF THE DIFFICULTY INDEX OF OBJECTIVE-TYPE SPELLING ITEMS .....	319
<i>Heston, Joseph C.</i>	
THE GRADUATE RECORD EXAMINATION VS. OTHER MEASURES OF APTITUDE AND ACHIEVEMENT .....	618
<i>Heston, Joseph C.</i>	
THE VALUES OF AN EDUCATIONAL GUIDANCE CLINIC .....	299

<i>Huppock, Robert</i>	
A REVIEW OF NEW BOOKS AND PAMPHLETS ON OCCUPATIONS FOR COLLEGE STUDENTS .....	485
<i>Hurrocks, John E. (with Maurice E. Troyer)</i>	
CASE STUDY TESTS OF ABILITY TO USE KNOWLEDGE OF HUMAN GROWTH AND DEVELOPMENT .....	23
<i>Jones, Edward S.</i>	
GRADATIONS OF DIRECTIVENESS IN COUNSELING .....	559
<i>Jones, Lonzo</i>	
FACULTY COUNSELING FOR FRESHMEN .....	564
<i>Keller, Robert J. (with Clifford P. Froehlich)</i>	
A TECHNIQUE FOR OBTAINING SUMMARY DATA FROM AG- GREGATE WEIGHTING SHEETS ON THE GRAPHIC ITEM COUNTER ATTACHMENT TO THE TEST SCORING MACHINE ..	429
<i>Kimber, J. A. Morris</i>	
THE INSIGHT OF COLLEGE STUDENTS INTO THE ITEMS ON A PERSONALITY TEST .....	411
<i>Kohn, Jr., Nathan</i>	
TRENDS AND DEVELOPMENT OF THE VOCATIONAL AND OTHER INTERESTS OF VETERANS AT WASHINGTON UNIVERSITY ....	631
<i>La Barre, Corinne</i>	
GRADUATE TRAINING FOR EDUCATIONAL PERSONNEL WORK .	538
<i>Layton, Warren K.</i>	
COMMUNITY RELATIONSHIPS IN OUT-OF-SCHOOL COUNSELING	127
<i>Leeds, Carroll H. (with Walter W. Cook)</i>	
MEASURING THE TEACHING PERSONALITY .....	399
<i>Lemmon, Vernon (with Philip H. DuBois)</i>	
MULTIPLE FORMS OF OBJECTIVE TESTS: A TEST SCORING TECHNIQUE .....	783
<i>Loeschner, Frank S.</i>	
THE COLLEGE COMMUNITY AND OCCUPATIONAL INTEGRATION OF MINORITIES .....	553
<i>Long, Louis (with James D. Perry)</i>	
ENTRANCE EXAMINATIONS AT THE CITY COLLEGE OF NEW YORK .....	765
<i>Mandell, Milton M.</i>	
THE SELECTION OF FOREMEN .....	385
<i>Morris, Harold M.</i>	
WHAT THE VETERAN THINKS OF EDUCATION .....	512
<i>Mosier, Charles I.</i>	
A CRITICAL EXAMINATION OF THE CONCEPTS OF FACE VALIDITY .....	191
<i>Newland, T. Ernest</i>	
CADET PERSONNEL PROBLEMS AND PROCEDURES AT THE UNITED STATES MILITARY ACADEMY .....	489
<i>Owens, William A.</i>	
AN EMPIRICAL STUDY OF THE RELATIONSHIP BETWEEN ITEM VALIDITY AND INTERNAL CONSISTENCY .....	281
<i>Pepinsky, Harold B.</i>	
APPLICATION OF INFORMAL PROJECTIVE METHODS IN THE COUNSELING INTERVIEW .....	135

<i>Pepinsky, Harold B.</i>	
MEASURING OUTCOMES OF CLASSROOM THERAPY .....	713
<i>Perry, James D. (with Louis Long)</i>	
ENTRANCE EXAMINATIONS AT THE CITY COLLEGE OF NEW YORK .....	765
<i>Pierson, George A.</i>	
SCHOOL MARKS AND SUCCESS IN ENGINEERING .....	612
<i>Powell, Norman J. (with William Brody)</i>	
A NEW APPROACH TO ORAL TESTING .....	289
<i>Reilly, Jean Waid (with Francis P. Robinson)</i>	
STUDIES OF POPULARITY IN COLLEGE: I. CAN POPULARITY OF FRESHMEN BE PREDICTED? .....	67
<i>Reilly, Jean Waid (with Francis P. Robinson)</i>	
STUDIES OF POPULARITY IN COLLEGE: II. DO DORMITORY ARRANGEMENTS AFFECT POPULARITY? .....	327
<i>Robinson, Francis P. (with Jean Waid Reilly)</i>	
STUDIES OF POPULARITY IN COLLEGE: I. CAN POPULARITY OF FRESHMEN BE PREDICTED? .....	67
<i>Robinson, Francis P. (with Jean Waid Reilly)</i>	
STUDIES OF POPULARITY IN COLLEGE: II. DO DORMITORY ARRANGEMENTS AFFECT POPULARITY? .....	327
<i>Rohrer, J. H.</i>	
FUTURE ENROLLMENTS IN U. S. INSTITUTIONS OF HIGHER EDUCATION .....	547
<i>Schneider, Gwendolen</i>	
THE ROLE OF THE PSYCHOLOGIST IN THE COUNSELING PROGRAM OF THE VETERANS ADMINISTRATION .....	117
<i>Seashore, Harold G.</i>	
VALIDATION OF THE STUDY OF VALUES FOR TWO VOCATIONAL GROUPS AT THE COLLEGE LEVEL .....	757
<i>Segel, David</i>	
THE VALIDITY OF A MULTIPLE APTITUDE TEST AT THE SECONDARY SCHOOL LEVEL .....	695
<i>Simpson, Ray H.</i>	
THE CRITICAL INTERPRETATION OF TEST RESULTS IN A SCHOOL SYSTEM .....	61
<i>Steinmetz, Harry C.</i>	
SELECTING PERSONNEL WORKERS .....	37
<i>Stephenson, Margaret</i>	
CURRENT TRENDS IN PERSONNEL WORK AT PRATT INSTITUTE	657
<i>Stewart, Naomi</i>	
RELATIONSHIP BETWEEN MILITARY OCCUPATIONAL SPECIALTY AND ARMY GENERAL CLASSIFICATION TEST STANDARD SCORE .....	677
<i>Strang, Ruth</i>	
THE COLLEGE PERSONNEL WORKER'S RESPONSIBILITY FOR THE IMPROVEMENT OF READING .....	603
<i>Stuit, Dewey B.</i>	
THE EFFECT OF THE NATURE OF THE CRITERION UPON THE VALIDITY OF APTITUDE TESTS .....	671

<i>Super, Donald E.</i>	
THE VALIDITY OF STANDARD AND CUSTOM-BUILT PERSONALITY INVENTORIES IN A PILOT SELECTION PROGRAM .....	735
<i>Super, Donald E.</i>	
VOCATIONAL INTERESTS AND VOCATIONAL CHOICE: PRESENT KNOWLEDGE AND FUTURE RESEARCH IN THEIR RELATIONSHIPS .....	375
<i>Taylor, E. K. (with R. H. Gaylord)</i>	
TABLE FOR USE IN THE COMPUTATION OF STATISTICS OF DICHOTOMOUS AND TRUNCATED DISTRIBUTIONS .....	441
<i>Thompson, Claude Edward</i>	
SELECTING EXECUTIVES BY PSYCHOLOGICAL TESTS .....	773
<i>Thompson, Florence</i>	
THE USE OF DORMITORIES FOR SOCIAL EDUCATION .....	648
<i>Triggs, Frances Oralind</i>	
THE DIAGNOSIS OF READING DEFICIENCIES AS AN AID TO REMEDIAL WORK .....	638
<i>Troyer, Maurice E. (with John E. Horrocks)</i>	
CASE STUDY TESTS OF ABILITY TO USE KNOWLEDGE OF HUMAN GROWTH AND DEVELOPMENT .....	23
<i>Van Dusen, A. G.</i>	
IMPORTANCE OF CRITERIA IN SELECTION AND TRAINING ....	(498)
<i>Vaughn, Kenneth W.</i>	
THE GRADUATE RECORD EXAMINATIONS .....	745
<i>Williamson, E. G.</i>	
COUNSELING AND THE MINNESOTA POINT OF VIEW .....	141
<i>Williamson, E. G.</i>	
IMPRESSIONS OF STUDENT PERSONNEL WORK IN GERMAN UNIVERSITIES AND IMPLICATIONS FOR AMERICA .....	519
<i>Worcester, D. A.</i>	
A MISUSE OF GROUP TESTS OF INTELLIGENCE IN THE SCHOOL	779
<i>Wrenn, C. Gilbert (with Robert Callis)</i>	
THE GED TESTS AS PREDICTORS OF SCHOLASTIC SUCCESS ..	93
<i>Wrenn, C. Gilbert</i>	
STUDENT PERSONNEL WORK IN A NATIONAL SETTING .....	707

STATEMENT OF THE OWNERSHIP, MANAGEMENT, CIRCULATION, ETC.,  
REQUIRED BY THE ACTS OF CONGRESS OF MARCH 3, 1907, AND MARCH  
3, 1933, OF EDUCATIONAL AND PUBLISHED  
quarterly at Lancaster, Pennsylvania  
District of Columbia.

Before me, a Notary Public in and for the State and county aforesaid personally  
appeared Frederic Kuder, who, having been duly sworn according to law, depose  
and says that he is the Editor of the EDUCATIONAL AND PUBLISHED  
MEASUREMENT, and that the following is, to the best of his knowledge and belief,  
a true statement of the ownership, management (and if a daily, weekly, and tri-weekly,  
or tri-weekly newspaper, the circulation), etc., of the aforesaid publication for the  
date shown in the above caption, required by the Act of August 24, 1912, as amended  
by the acts of March 3, 1907, and July 2, 1948 (section 557, Postal Laws and Regu-  
lations) printed on the reverse of this form, to wit:

1. That the names and addresses of the publisher, editor, managing editor, and  
business managers are  
Kuder, 917 Fifteenth St., N. W., Wash-  
ington, D. C. Editor,  
Managing Editor, Fred.  
Fifteenth St., N. W., D. C.  
Business Manager, Maxine M. Lyne, 917 Fifteenth St., N. W., D. C.

2. That the owner is: (If owned by a corporation, its name and address must be  
stated and also immediately thereunder the names and addresses of stockholders own-  
ing or holding one per cent or more of total amount of stock. If not owned by a  
corporation, the names and addresses of the individual owners must be given. If  
owned by a firm, company, or other unincorporated concerns, its name and address,  
as well as those of each individual member, must be given.) Frederic Kuder, 917  
Fifteenth Street, N. W., Washington, D. C.

3. That the known bondholders, mortgages, and other security holders owning  
or holding 1 per cent or more of total amount of bonds, mortgages, or other securities  
are: (If there are none, so state) None.

4. That the two paragraphs next above, giving the names of the owners, stock-  
holders, and security holders, if any, contain not only the list of stockholders and  
security holders as they appear upon the books of the company but also, in cases  
where the stockholder or security holder appears upon the books of the company as  
trustee or in any other fiduciary relation, the name of the person or corporation for  
whom such trustee is acting is given; also that the said two paragraphs contain  
statements embracing affiant's full knowledge and belief as to the circumstances and  
conditions under which stockholders and security holders who do not appear upon the  
books of the company as trustees, hold stock and securities in a capacity other than  
that of a bona fide owner; and this affiant has no reason to believe that any other  
person, association, or corporation has any interest direct or indirect in the said  
stock, bonds, or other securities than as so stated by him.

5. That the average number of copies of each issue of this publication sold or  
distributed, through the mails or otherwise, to paid subscribers during the twelve  
months preceding the date shown above is: not a daily. (This information is re-  
quired from daily, weekly, semi-weekly, and tri-weekly newspapers.)

Signed: Frederic Kuder, Editor. Sworn to and subscribed before me this 30th  
day of September, 1947. Patrick H. McCormick, Notary Public, D.C.  
(Seal) (My commission expires July 14, 1948.)

EDUCATIONAL and  
PSYCHOLOGICAL



MEASUREMENT

**VOLUME SEVEN, NUMBER ONE, SPRING**

<i>Technical Aspects of the Fourth Annual Science Talent Search.</i> HAROLD A. EDGERTON and STEUART HENDERSON BRITT .....	3
<i>Case Study Tests of Ability to Use Knowledge of Human Growth and Development.</i> JOHN E. HORROCKS and MAURICE E. TROYER .....	23
<i>Selecting Personnel Workers.</i> HARRY C. STEINMETZ .....	37
<i>What Constitutes a Minimal School Testing Program.</i> WALTER N. DUROST .....	45
<i>The Critical Interpretation of Test Results in a School System.</i> RAY H. SIMPSON .....	61
<i>Studies of Popularity in College: I. Can Popularity of Freshmen Be Predicted?</i> JEAN WAID REILLY and FRANCIS P. ROBINSON .....	67
<i>Measurement Abstracts</i> .....	73
<i>The Contributors</i> .....	85
<i>The American College Personnel Association Supplement</i> .....	89

## CORRECTION

The following paragraph and footnote were inadvertently omitted at the end of the article, "New Standards for Test Evaluation" by J. P. Guilford in the Winter, 1946, issue of EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT:

As a complement to the presentation of factor loadings of tests, we need parallel information concerning the factorial compositions of jobs. This is not the sole responsibility of the test maker. The accumulation of such information about jobs is a project in itself, and I propose it as one of major social importance. From the two sources of information, factor saturations in tests and factor saturations in jobs, validities of tests and of combinations of tests can be fairly closely predicted. These two sources of information carry us to the heart of the test-evaluation problem.<sup>8</sup>

---

<sup>8</sup> This discussion has been confined to the practical use of tests in selection, classification, and guidance. The evaluation of tests for research purposes is still a different problem that cries for attention. Many an investigation using tests as measuring instruments is almost entirely worthless because complex, ambiguous tests were used. Unless tests measure single functions, full experimental control is not achieved. This applies to studies of comparisons of groups, functional deficit, and of heredity and environment.



# TECHNICAL ASPECTS OF THE FOURTH ANNUAL SCIENCE TALENT SEARCH

HAROLD A. EDGERTON  
Ohio State University  
AND  
STUART HENDERSON BRITT  
McCann-Erickson, Inc.

## I. *Background*

THE present article is a statistical review of certain aspects of the selection procedures used in the Fourth Annual Science Talent Search.<sup>1</sup> This annual competition is conducted jointly by Science Clubs of America, administered by Science Service, and the Westinghouse Electric Corporation.

The selection of students to be trained as scientists involves, first of all, a description of the kinds of persons to be selected. This description must then be made operational in terms of the development of appropriate instruments and procedures so that the kinds of persons demanded by the description will be selected.

In this connection, the selection procedures for the Fourth Annual Science Talent Search were as follows:

Of about 15,000 entrants, complete entry materials—namely, *Science Aptitude Examination*, personal data, scholarship record, and scientific essay—were received from 2,746 high-school seniors (1,970 boys and 776 girls). This group of students, then, were considered to have completed the first hurdle.

The second hurdle, the *Science Aptitude Examination*, was similar to that of the previous year. Instead of two parts, however, it contained three: the first third was composed of scientific problems, with multiple-choice answers; the second part consisted of a paragraph reading test on materials from

---

<sup>1</sup> Compare Edgerton, H. A. and Britt, S. H., "The First Annual Science Talent Search," *American Scientist*, XXXI (1943), 55-68.

various fields of science; and the last third was made up of questions reflecting judgment and breadth of scientific reading. Scores on the examination constituted the second hurdle, which reduced the number of contestants from 2,746 to 1,082. Of these, 787 were boys and 295 were girls, the proportions being the same as the ratio of boys to girls with complete entrance materials.

The third hurdle was based on the academic record of the individual; the high-school record "composite" score was the sum of the score for relative rank in the high-school class and units of high-school science taken, weighted 5: 1 respectively. The 611 highest (440 boys and 171 girls) were deemed to have passed this hurdle.

The fourth step was an evaluation of the recommendations made by high-school faculty members. Trained raters scored this information in terms of specific actual accomplishments; and on this basis the population was then reduced to 215 boys and 85 girls—containing the 40 trip winners and the 260 students who were given Honorable Mention.

The essays of these 300 were read separately and scored by three members of the staff of Science Service. Each contestant had written an essay of about 1,000 words on the subject, "My Scientific Project," telling what he or she was doing or planned to do in science in the way of experimentation or other research activity.

At this point, on the basis of all the evidence thus far accumulated—the scores on the *Science Aptitude Examinations*, high-school records, recommendations, and essays—the judges made the selection of the 40 trip winners (29 boys and 11 girls) to the Science Talent Institute held in Washington, D. C. The names and geographical localities represented were completely unknown throughout, for this information had been blanked out so that identification was by serial number only. No questions concerning either race or religion appeared in any of the entry materials.

The selections at Washington in March, 1945, of the Westinghouse Science Scholarship winners were made by the present writers in cooperation with Dr. Harlow Shapley. The decisions

were based on the "over-all" previous evidence, plus information obtained from individual, standardized 15-minute interviews by each of the three judges. The interviews were designed to determine how well the contestant seemed fitted for a promising career in science, and were supplemented by data obtained in additional interviews by a psychiatrist and a psychiatric social worker.

## II. *Reliability of Examination*

The *Science Aptitude Examination* used in the Fourth Annual Science Talent Search furnished part of the evidence used to select boys and girls who were above average in intelligence, and who seemed capable of learning the techniques and subject matter of sciences in colleges and universities. The examination was intended to discriminate among those having the greater potentialities as scientists. It was designed as an academic aptitude test, using science materials, concepts, ideas, facts, and vocabulary, to estimate potentiality for learning about science rather than current comprehension of scientific subject matter.

There were three parts to the test. Part A contained fifty items arranged in four-answer multiple choice form. These items were taken from many fields of science and were deemed to reflect general breadth of interest in various areas of science, and ability to deal with scientific materials effectively. The estimated reliability coefficient for this part is .63 for boys and .61 for girls (Kuder-Richardson, Case IV).<sup>2</sup> Several sample items are reproduced below:

8. The word *sidereal* is most likely to be used by

- ( ) 1. an astronomer
- ( ) 2. an electrical engineer
- ( ) 3. an actuary
- ( ) 4. a chemist

9. Which of the following is not a primate?

- ( ) 1. orangutan
- ( ) 2. gibbon
- ( ) 3. lemur
- ( ) 4. lemming

---

<sup>2</sup> Kuder, G. F. and Richardson, M. W., "The Theory of the Estimation of Test Reliability," *Psychometrika*, II (1937), 151-160.

11. A kilogram (1,000 grams) is equal to 2.2046 pounds. There are 15,432.356 grains in a kilogram. How many grains are there in an ounce?

( ) 1.  $\frac{1,000}{2.2046} \times \frac{16}{15,432.356}$

( ) 2.  $\frac{15,432.356}{(16)(2.2046)}$

( ) 3.  $\frac{15,432.356}{2.2046} \times 16$

( ) 4.  $\frac{(16)(2.2046)}{15,432.356}$

25. A certain tree adds one ring each year. Its rings are  $\frac{1}{8}$  inch thick on the average. In 1937 the tree, 3 feet above the ground, was 4 inches in diameter. What should one expect its diameter to be in 1946?

( ) 1.  $4\frac{1}{8}$  inches

( ) 2.  $5\frac{1}{8}$  inches

( ) 3.  $5\frac{3}{8}$  inches

( ) 4.  $6\frac{1}{8}$  inches

29. Which of the four words belongs least with the other three?

( ) 1. quiescent

( ) 2. dormant

( ) 3. latent

( ) 4. inert

Part B was a paragraph reading test using paragraphs drawn from various areas of science: physics, bacteriology, logic, engineering, chemistry, genetics, physiology. Following each paragraph were several questions in four-answer multiple choice form which could be answered by inference or by reasoning from the information given in the paragraph. Here is an example:

PARAGRAPH B: Disregarding the resistance of the medium through which it falls, a body falls 16 feet in the first second of its descent, during the second second,  $16 + 32 = 48$  feet; third second,  $16 + 64 = 80$  feet; fourth second,  $16 + 96 = 112$  feet; fifth second,  $16 + 128 = 144$  feet; etc. The total distance fallen by a body at the end of the  $n^{\text{th}}$  second is given in feet by multiplying the square of the time in seconds by 16. Thus, at the end of the first second it has fallen 16 feet; at the end of the second second,  $2 \times 2 \times 16 = 64$ ; at the end of the third second,  $3 \times 3 \times 16 = 144$  feet; at the end of the fifth,  $5 \times 5 \times 16 = 400$  feet.

## QUESTIONS FOR PARAGRAPH B

56. If  $t$  = time in seconds, the formula expressing the distance a ball of iron will fall during the  $t^{\text{th}}$  second after it is released is

- ( ) 1.  $16 + 32(t - 1)$   
 ( ) 2.  $16 + 32t$   
 ( ) 3.  $16 + 32(t + 1)$   
 ( ) 4.  $16 + 32t^2$

58. How far will a body fall during the third half-second of its fall?

- ( ) 1.  $16 + \frac{48}{2}$   
 ( ) 2.  $(16)(1.5)^2 - 16$   
 ( ) 3.  $16 \times \frac{3}{2}$   
 ( ) 4.  $16 + 32 \times \frac{3}{2}$

59. A bullet was shot vertically to an altitude of 765 yards. Assuming that it requires as long for its ascent as for its descent, about how long was the bullet in flight?

- ( ) 1.  $11\frac{1}{2}$  seconds  
 ( ) 2. 13 seconds  
 ( ) 3.  $22\frac{1}{2}$  seconds  
 ( ) 4. 26 seconds

Part C involved a greater variety of response by the contestants. It was used for experimental purposes only, and no score on Part C was actually used in the selection process. Thirty items were devoted to measuring skill in drawing inferences from data. Five were concerned with such questions as "Why is the sun hot?" "Why does hot air rise?" etc. The next five questions asked the contestant to name five living scientists and to state in less than ten words the outstanding achievement of each one. The remaining ten questions were ones to which brief free answers could be given.

As indicated above, the high-school records, recommendations of high-school teachers, and ratings of a scientific essay, were used in the order named as successive hurdles in selecting the winners.

For purposes of the present analysis of the statistical aspects of the selection procedures two groups were studied:

1. A random sample was drawn by including every tenth contestant who had completed all three of the forms required. As the packages of materials from each high school were opened each contestant was assigned a serial number and this serial number was placed on each of the forms for each contestant. Every serial number ending in 7 (the units digit) was chosen.

2. The final 300 selected on the basis of the *Science Aptitude Examinations*, high-school records, and the recommendations were the second group selected for study. The 40 trip winners were selected from this group on the basis of their essays and the other accumulated information; and each of the remaining 260 were awarded an Honorable Mention. Each of the 300 who had a serial number with seven as the digit was included in both samples.

Table 1 shows the distribution of scores (Part A plus Part B) for boys and for girls. Based on these distributions the estimated reliabilities of the total score for Parts A and B together are .78 for boys, and .74 for girls.

The reliability coefficient of the Aptitude Examination and of its parts was estimated by the Kuder-Richardson formula, Case IV. The resulting values are underestimates of the reliabilities, since the test data do not closely fit the assumptions made by the formula; namely, that the intercorrelations of the items are equal, and that the items are of equal difficulty.

These distributions show the type of skewness desired. The typical difference in scores for boys and girls—favoring the boys—is also illustrated.<sup>3</sup> Table 2 shows the distribution of scores for Parts A, B, and C for boys and girls derived from the random sample. This, too, illustrates the higher scores of the boys as compared with the girls.

It is assumed, from the skewness of the frequency distribution of scores and from a study of the difficulties of such items, that the reliability of the test is highest in the upper half of the score ranges. This is the region in which all of the discriminations based on the test scores are made. About the highest one-third of the test scores are assumed to have cleared

<sup>3</sup> Edgerton, H. A. and Britt, S. H., "Sex Differences in the Science Talent Test," *Science*, C (1944), 192-193.

the first hurdle. The remaining two-thirds did not survive this hurdle and are not considered further as contestants. Test scores within this restricted range are again considered in selecting the 40 trip winners from the 300. The reliability of Part C as a whole probably is lower than Parts A and B, due to the subjectivity of scoring of the items 131 to 150.

For boys the correlations of the several parts were as follows: A and B, 0.53; A and C, 0.48; and B and C, 0.60. For girls the correlations were 0.57, 0.70, and 0.51 respectively. In other words, there are fair correlations among Parts A, B, and C, but they are low enough to suggest that to some extent each part of the Examination measured something different.

### III. Item Analysis

A study of the items of Parts A and B of the *Fourth Annual Science Aptitude Examination* was made. Tabulations of the responses marked for each item were obtained for three groups: those with high scores, those with above average scores, and average to slightly below average scores. Since the Examination was designed to discriminate within the above average groups, such groupings for the study of item discrimination give more useful information than the use of extremely high versus extremely low scoring papers. The three groups (scored for Parts A and B) were: Group A—scores 54 and higher; Group B—scores 45 to 53; Group C—scores 35 to 44. (See Table 1).

The proportion of boys marking each alternative to each item was obtained for each of the three groups, using the random sample of all contestants. An index of goodness of each item was produced by considering only the correct response. For the correct response to each item, the index for the item was obtained by taking twice the difference of the per cents in Groups A and B who answered the item correctly plus the difference of the per cents of Groups B and C who answered the item correctly. That is,  $\text{Index of Goodness} = 2(A - B) + (B - C)$ . Such an index gives a premium to that item which discriminates at the higher-score levels. It also gives some advantage to those items which are rather difficult.

TABLE 1

*Distributions of Total Scores on the Science Aptitude Examination  
(Part A Plus Part B) for Boys and Girls*

Examination Score	Boys		Girls	
	All	Random	All	Random
80-85	2	1	..	..
75-79	13	2	..	..
70-74	21	1	1	..
65-69	55	6	4	1
60-64	116	8	15	2
55-59	200	21	33	3
50-54	324	26	74	7
45-49	364	32	127	14
40-44	357	34	152	8
35-39	267	24	159	13
30-34	170	19	114	14
25-29	51	12	63	10
20-24	24	1	26	..
15-19	5	1	6	2
10-14	1	...	1	..
5-9	...	..	1	..
Total ... ..	1970	188	776	74

The distribution of the indexes for the 100 items of Parts A and B for boys is shown in Table 3. For purposes of study, the 100 items were divided into 3 groups: the 29 items with

TABLE 2

*Distribution of Scores of Boys and Girls for Each of the Parts of the Science  
Aptitude Examination (Highest Possible Score in Each Part—50)*

Score	Boys			Girls		
	A	B	C	A	B	C
45-47	1	...	...	..	.	..
42-44	3	.	.	..	..	..
39-41	3	2	...	1	..	.
36-38	1	...	2	0	.	..
33-35	17	4	8	3	..	1
30-32	23	5	20	5	1	3
27-29	38	13	31	6	3	2
24-26	30	32	38	17	10	5
21-23	32	23	31	15	8	19
18-20	28	38	22	13	15	12
15-17	9	30	17	10	12	9
12-14	3	21	8	4	13	10
9-11	.	14	7	..	7	8
6-8	..	6	5	..	2	5
3-5	..	.	..	..	3	..
Total . . .	188	188	188	74	74	74



the highest indexes were classed as "Good," the 42 with medium indexes as "Average," and the 29 lowest as "Poor." Those with indexes labeled "Average" appear to be quite acceptable items. It must be recognized, of course, that the goodness of an item here refers to the extent to which the item makes the same kind of discrimination as does the examination as a whole. A check was made to see if some kinds of items appeared to be

TABLE 3  
*Distribution of Indexes of Goodness of Items for Boys*

Indexes	Part A	Part B	Total
85-9	1	1	2
80-4	1	1	2
75-9	..	..	..
70-4	1	2	3
65-9	3	1	4
60-4	..	..	..
55-9	2	2	4
50-4	3	4	7
45-9	3	4	7
40-4	1	3	4
35-9	4	2	6
30-4	9	5	14
25-9	5	6	11
20-4	4	3	7
15-9	3	3	6
10-4	4	3	7
5-9	..	2	2
0-4	4	2	6
- 1- - 5	..	3	3
- 6- - 10	1	1	2
- 11- - 15	1	..	1
- 16- - 20	..	2	2
Total .....	50	50	100

better than others. Within any class of items, e.g., vocabulary, arithmetic, use of formulas, general information, mechanical comprehension, etc., both good and poor items appeared. Items which were too difficult or too easy were not good discriminators. However, had the critical ratio been used instead of the item index, the picture with reference to the more difficult items apparently would have appeared slightly more optimistic, since differences between proportions near 100% or 0% have smaller sampling errors than those differences based on proportions near 50%.

It has been observed that there are differences in the mean

scores made by boys and girls on the Examination. Such differences could be due to the use of more items which were favorable to boys than to girls. To investigate this possibility, the item responses of a sample of girls of the B and C score range were tabulated in the same way as for the boys. If the difference of the percentages of girls and boys (B and C groups) answering the item correctly was greater than 10, the item was assumed to have a sex bias. This index is independent of the magnitude of the total score. In Part A, 22 of the 50 items showed a sex bias; in Part B, only 12 of the 50 items showed

TABLE 4  
*Intercorrelations of Selectors*  
(Boys, Random Sample)

	N = 152			
	Sc. Apt. Exam.	H.-S. Record	Recommen- dation	Honors
Science Aptitude Examination ...	..	.33	.26	.28
High-School Record . . . . .	.33	..	.18	.42
Recommendation . . . . .	.26	.18	..	.31

such a bias. The bias in Part B favors the girls (10 of the 12 items having such bias), while in Part A it favors the boys (14 of the 22). This suggests that some of the sex bias may be a function of the kind of items included, but it is difficult to draw any systematic picture of sex difference for items so far as this test is concerned, except that the items involving pulleys and gears are a bit favorable to the boys. For practically every item where a sex bias was found, a similar type of item was found with an opposite sex bias.

#### IV. *Intercorrelations of Selectors*

Table 4 shows the intercorrelations of the selectors in the random sample for boys, and Table 5, for girls. The *Science Aptitude Examination* score is the total for Parts A and B. The high-school record is the net relative standing in the high-school class, expressed in terms of the abscissa of the normal curve. The criterion, Honors, was valued 1 if the contestant won honors (winner or honorable mention), otherwise a 0. The selectors show relatively low intercorrelations, suggesting

TABLE 5  
*Intercorrelations of Selectors*  
*(Girls, Random Sample)*

	<i>N</i> = 64			
	Sc. Apt. Exam.	H.-S. Record	Recommen- dation	Honors
Science Aptitude Examination ...	..	.36	.36	.50
High-School Record .....	.36	..	.23	.39
Recommendation ..	.36	.23	..	.37

that they are fairly independent of each other. The correlations of the selectors with the criterion are not high. This is due in large part to the use of the successive hurdles method rather than a weighted aggregate of scores.

The next question is, just what correlation each of the selectors bore to the selection of the 40 winners from the 300. The correlations of total Examination score (based on both Parts A and B), high-school relative standing, recommendation score, and aggregate essay rating with each other, and also with the selection, are shown in Tables 6 (boys) and 7 (girls). For purposes of the analysis, the 40 trip winners were scored 1, and the 260 honorable mentions 0. The intercorrelations of the selectors are about the same for both boys and girls, but distinctly lower than shown in Tables 4 and 5 for the random sample. This is due to the inclusion in the 300 of only those who had fairly high test scores, those who were in the top 6 per cent of the high-school class, and those who had high recom-

TABLE 6  
*Intercorrelation of Selectors for Boys*

	<i>N</i> = 215					
	Sc. Apt. Exam.	H.-S. Record	Recommen- dation Score	Essay Rating	Contest Stand- ing	Partial Regress- ion Coeffi- cients
Science Aptitude Exam- ination .....	..	.08	.05	.08	.44	.40
High-School Record ...	.08	..	.23	.05	.01	.00
Recommendation Score ..	.05	.23	..	.14	.24	.16
Essay Rating .....	.08	.05	.14	..	.43	.38

mentation scores. The partial regression coefficients, using contest standing as a criterion, shown in Tables 6 and 7, indicate that the essay rating and Examination score carried practically all of the linear discrimination. This might well be expected, since most of the significant variance of the high-school relative standing and recommendation score was used in selecting the 300, leaving such a narrow range of talent, as reflected by these two indexes, that discrimination within the range is largely a matter of chance. The multiple correlation of the four selectors with contest standing is .61 for boys and

TABLE 7  
*Intercorrelation of Selectors for Girls*

	<i>N</i> = 85					
	Sc. Apt. Exam	H.-S. Record	Recom- menda- tion Score	Essay Rating	Contest Stand- ing	Partial Regres- sion Coeffi- cients
Science Aptitude Exami- nation . . . . .	.	.12	.01	.30	.52	.43
High-School Record . .	.12	..	.03	.05	.06	.02
Recommendation Score .	.01	.03	..	.12	.10	.07
Essay Rating . . . . .	.30	.05	.12	..	.43	.30

.60 for girls. The size of these correlations suggests that factors other than those reflected by the linear relationships could have been operative. These include such concepts as patterning of the predictor scores, re-evaluation of recommendations, consideration of personal data not directly used in scoring the recommendation, and the like.

### V. *Correlations of Essay Ratings*

Since the essay ratings were used in selecting the winners and were shown statistically to be of importance in this selection, a more detailed review of these should be reported. Each of the 300 essays was rated independently by three raters. The essays were identified only by serial number; names and addresses were covered by masking tape. Each essay was graded *Good*, *Fair*, or *Poor*. An aggregate rating was obtained by giving the values 3, 2, and 0, respectively, to the categories and

summing the scores for the three raters for each essay. In any case when one rater assigned a *Good* and another a *Poor* to the same essay, the raters were asked to discuss their basis of rating and perhaps to revise their ratings so that no such extreme discrepancy remained. In two cases, however, the raters were unable to come to such an agreement. The number of such changes was sufficiently small that they would not have changed any of the inter-rater correlations more than 0.03.

Tables 8 and 9 show the intercorrelations of the final ratings of the three raters, and the correlation of their ratings with the total rating and with the final selection.

TABLE 8  
*Correlations of Essay Ratings for Boys*

	N = 215				
	Rater A	Rater B	Rater C	Total Rating	Selection
Rater A .....	..	.60	.62	.83	.26
Rater B .....	.60	..	.66	.85	.40
Rater C .....	.62	.66	..	.87	.37
Total Rating .....	.83	.85	.87	..	.43
Selection .....	.26	.40	.37	.43	..

TABLE 9  
*Correlations of Essay Ratings for Girls*

	N = 85				
	Rater A	Rater B	Rater C	Total Rating	Selection
Rater A .....	..	.55	.54	.82	.32
Rater B .....	.55	..	.61	.85	.49
Rater C .....	.54	.61	..	.85	.32
Total Rating .....	.82	.85	.85	..	.44
Selection .....	.32	.49	.32	.44	..

## VI. *Additional Information from Interview*

Within the group selected as winners, a record was made of the order in which they were selected. This order was a function of the available information, essay ratings, *Science Aptitude Examination* scores, high-school records, and recommendations.

When the winners were interviewed a record was made of the rank order assigned on the basis of the interviews alone. Within this highly select group the rank difference coefficients of correlation for these two rankings were 0.12 for the 29 boys, and 0.35 for the 11 girls. This points sharply to the fact that within the very narrow range of scores of the winners the interview brings in new and independent information for use in the selection of scholarship winners.

#### VII. *Examination Scores Independent of Courses in Science and Mathematics*

It is possible that scores on the *Science Aptitude Examination* are in part a function of the amounts of science and mathematics studied in high school. In constructing the Examination, it was hoped that it could be made fairly independent of such influences, letting it reflect aptitude for college study, and in general the ability to understand, to think, and to reason in terms of the concepts and vocabulary of the sciences.

Were the Examination scores independent of the number of courses taken in science and mathematics?

Two null hypotheses were stated as the basis for answering the question:

1. The amount of mathematics taken by winners in high school bears no significant relationship to the total scores on the test.
2. The amount of science taken by winners in high school bears no significant relationship to the total scores on the test.

1. To investigate the first null hypothesis, the test scores of the 300 who won honors were analyzed with respect to amounts of mathematics taken in high school. The amounts of mathematics taken by the 300 in the honors group are shown in Table 10. The  $F$  ratio for boys (ratio of mean square of Examination scores for amounts of mathematics to the mean square for error) was found to be  $2.73/2.25 = 1.21$ , with five degrees of freedom for the numerator and 209 for the denominator. (The 1% value for  $F$  is 3.11.) For girls, the  $F$  ratio was  $1.16/1.72 = 0.67$ , with 4 degrees of freedom in the numerator and 80 in the

TABLE 10

*Half-Units of Mathematics Taken in High School by the Honors Group*

Half-Units of Mathematics Taken	Boys	Girls
9 or more	7	15
8	75	19
7	51	36
6	59	4
5	15	11
4 or less	8	85
Total	215	

denominator. This shows that the first null hypothesis was not proven untenable, and hence will be allowed to stand as true. It must be noted that the statistical method used cannot prove the hypothesis true. In view of the fact that the hypothesis was not shown to be untenable, it is assumed that it is valid.

2. With respect to the second null hypothesis, the test scores were investigated relative to the number of different sciences taken in high school. Table 11 shows the numbers of sciences taken in high school by the honors group. The  $F$  ratio for boys was  $5.78/1.71 = 3.38$  with 4 degrees of freedom for the numerator and 210 for the denominator. This ratio is only slightly less than the 1% value, 3.41. For the girls, the  $F$  ratio was  $1.89/1.63 = 1.16$  with 4 and 80 degrees of freedom for the numerator and denominator, respectively. The ratio for boys is near the borderline of significance. The ratio for the girls is non-significant. Hence, the second null hypothesis is deemed to be essentially true, not having been shown clearly to be untenable.

TABLE 11

*Number of Sciences Taken in High School by the Honors Group*

Number of Sciences Taken	Boys	Girls
5 or more	7	2
4	86	13
3	75	36
2	39	33
1	8	4
Total	215	85

It is possible, of course, that an analysis of only the scores of the 300 who received honors restricted the range of the scores, so that significant variance was covered up. Accordingly, a similar analysis was made from the random sample of the contestants; the sample was drawn from winners, honorable mentions, and non-winners. The analysis was made for Parts A and B separately. Table 12 shows the  $F$  ratios for the sev-

TABLE 12  
*F Ratios for Science Aptitude Examination Scores for Boys and for Girls  
for Amounts of Mathematics and Amounts of Science  
Taken in High School*

	Part A		Part B	
	Boys	Girls	Boys	Girls
Science				
$F$ .....	*	*	2.39	*
1% level .....	2.94	5.01	2.94	5.01
Mathematics				
$F$ .....	1.95	2.81	2.77	1.67
1% level .....	3.45	5.01	3.45	5.01

\* Less than unity.

eral comparisons. For different amounts of mathematics and for different amounts of science taken in high school there were no significant differences in Examination scores. This demonstrates that the test scores are fairly independent of the amounts of training in both mathematics and science for the contest group. It is still possible that if the test were given to a random sample of high-school pupils, not just those interested in science and entering the Science Talent Search, the effects of amounts of mathematics and science taken might be reflected in the variance of the test scores.

### VIII. *Inference Questions and Knowledge of Students*

Part C of the *Fourth Science Aptitude Examination* was designed primarily to investigate the possibilities of using other examination approaches to the selection of science talent. As already indicated, Part C may be thought of as made up of three kinds of questions: questions 101-130, designed to measure skill in identifying inferences made from paragraph mate-



rials; brief answer questions; questions such as those asking for the names of five living scientists and their outstanding achievements.

The inference questions were in many ways similar to the questions in Part B of the examination. The questions, particularly in Part B, force the contestant to draw inferences, but without being so labelled.

The scores on the subjective questions are probably somewhat unreliable since they depended largely on the judgment of the scorer. They did furnish information and suggestions which were useful in the construction of test materials for the Fifth Examination. Also, the cost of scoring such kinds of questions, both in terms of time and money, makes their use inadvisable, unless the number of ways of stating the correct answer can be reduced so that only a moment is required to judge their correctness, and so that that judgment can be made reliably by an accurate clerk.

The questions asking for the names and outstanding achievements of five living scientists yielded very interesting results. Listing in *American Men of Science* was used as a scoring guide for American scientists. For non-American scientists, those named were sufficiently outstanding that their identity and chief contributions could be readily verified by reference to recent science news or by checking with men in the particular subject-matter field.

In the random sample, 189 different names were mentioned, a total of 371 correct mentions and 224 incorrect mentions (i.e., not a scientist, not living, or incorrect identification of work). Einstein was mentioned correctly by 129 and incorrectly by 4. The second in number of mentions was Millikan with 29 correct and 1 incorrect. The most amusing identifications were Eve Curie—"developed radar," and Irving Langmuir—"improving the election." The number of names of scientists, now dead, was surprising. Of the 189 names mentioned, 20 were of persons not now living. These included such names as Lister, Alexander Graham Bell, Burbank, Diesel, Carver, Edison, and Charles M. Hall.

### IX. *Choices of Colleges and Technical Schools*

The 215 boys and 85 girls who won honors were widely distributed in their choice of colleges and technical schools. Of the 215 boys, 39 wanted to attend the Massachusetts Institute of Technology; 11 each, Harvard and Columbia; 7 each, the University of California, the University of Illinois, and the University of Wisconsin; 6 each, the University of Chicago and the Carnegie Institute of Technology. Sixty-nine other schools were mentioned, and 11 contestants mentioned no school.

Eighty-nine wished to study engineering, 35 physics, 33 chemistry, and 20 medicine. Such fields as biology, dentistry, bacteriology, mathematics, biochemistry, meteorology, and archaeology were mentioned.

Of the 85 girls, 7 wanted to attend the University of Michigan. Four each were interested in Cornell, the University of Pennsylvania, and the University of Wisconsin. Forty-four other schools were mentioned, and five contestants mentioned no school.

Fourteen girls wanted to study medicine, 12 chemistry, and 5 each physics and nursing. Other fields included medical technology, metallurgy, meteorology, medical research, food technology, aeronautics, genetics, etc. Only 2 were undecided as to field of study.

### X. *Summary*

A study of several aspects of the selection procedures used in the Fourth Annual Science Talent Search was based on evidence obtained from two samples: a random 10 per cent sample of all contestants who submitted complete credentials, and the 300 who won honors in the Search.

The estimated reliability of the *Science Aptitude Examination*, as a minimum, was 0.78 for boys and 0.74 for girls, and probably higher within the upper score levels where better discrimination is desired. The intercorrelations among the three parts of the Examination were moderate, averaging 0.58. No particular kind of item was superior to another as far as reflected by the item indexes used.

The mean score for boys was higher than for girls. This is

consistent with the findings of the three previous searches. In searching for kinds of items which would account for such differences, only those involving gears or pulleys showed any consistent bias.

Investigating, *a posteriori*, the relationships between the selectors and the criterion, winning honors or not winning honors, the correlations were fairly low, ranging from 0.28 to 0.50. This is to be expected since a successive hurdles procedure rather than an aggregate score technique was used.

In the *a posteriori* view of the selection of the winners from the 300, the Aptitude Examination and the Essay ratings carried the linear discrimination.

The reliability of essay ratings, based on average intercorrelations of three Essay raters is quite satisfactory. The average intercorrelations were 0.63 for boys' essays and 0.75 for girls' essays.

In building the Examination, the intent was to make it an academic aptitude test using science materials as a vehicle, rather than to make it a science knowledge examination. It was shown that the Examination scores are independent of the different amounts of mathematics taken by the contestants, and also that the scores are independent of the number of sciences studied in high school by the contestant.

Over half of the boys who won honors wanted to study engineering, chemistry, or physics. Twenty were looking toward medicine. The girls' choices of fields were not so concentrated in the physical sciences as were those for the boys; the field of medicine was the choice of 14 (16 per cent), while chemistry was the field chosen by 12 (14 per cent).



# CASE STUDY TESTS OF ABILITY TO USE KNOWLEDGE OF HUMAN GROWTH AND DEVELOPMENT

JOHN E. HORROCKS

Ohio State University

AND

MAURICE E. TROYER

Syracuse University

THE purpose of this paper is to describe the construction and use of three reasonably valid and reliable instruments developed as a means of measuring students' and teachers' ability to apply the facts and principles of human development.

Traditionally student progress is appraised by testing for the acquisition of facts and principles in the discipline being studied. The test may be objective or subjective. It may be purely memoriter or it may give opportunity for the student to recognize the truth or falsity of generalizations or to draw inferences. But essentially, such a test is not dynamic in that it presents isolated instances. It is artificial in that there is seldom any attempt to even simulate a life situation. It tends to test knowledge or recognition of facts and principles. Thus there is the implied assumption that ability to retain and repeat facts and principles is evidence that students can use them effectively.

Unfortunately, there is evidence to support the belief that the knowledge of facts and principles does not assure the ability to apply those facts and principles to a life situation. Horrocks (5) found that knowledge of facts and principles about adolescent behavior are positively but not highly related to the ability to make a diagnosis or to identify appropriate remedial procedures. This would also appear to be true in other areas. The situation is well summed up by Eurich and Cain (3) "tests of knowledge are not valid measures of teaching ability, nor

are tests of judgment in art adequate measures of ability to paint a picture." This is not, however, to say that knowledge of facts and principles is unimportant. Students cannot think in a vacuum.

In teaching the psychology of behavior, particularly in its applied aspects, it is essential that a specific effort be made to insure that students be able to apply the knowledge they have when they are faced with situations for which their training is ostensibly preparing them. It is well to be able to list the symptoms of deafness, a conflict situation, or a given educational deficiency. But it is also important to recognize such symptoms in another person with whom one comes in every-day or professional contact. It is desirable to be intellectually aware of the inter-relatedness of the home, the school, and the community in a child's welfare. But the ultimate and essential goal is to make a correct diagnosis and to help improve the environment for an adolescent with whose welfare one is concerned.

The three case study tests subsequently described are the results of a serious effort to improve appraisal of progress toward these important goals.

### *Preliminary Steps*

The construction of an instrument which would measure ability to use knowledge or facts about human development is not without its difficulties. The major difficulty which presents itself is one of practical expediency. Ability to use or to apply such knowledge might be measured by placing an individual in a counseling situation where he would have to work and apply his knowledge under the observation of the examiner. There are several difficulties, however. First is the difficulty of finding a series of diverse situations which would allow the individual to demonstrate in practice the usefulness of various phases of his knowledge. Second, if any experimental use is to be given to the test, is the difficulty of setting up situations similar enough so that comparable measures may be obtained for large numbers of subjects. Third is the difficulty of time and available facilities if the test is to have practical use.

The possibility of having students make case studies was considered, but was discarded since the construction and evaluation of student-made case studies, although useful for certain purposes, would have most of the disadvantages listed under the discussion of personal observation.

For purposes of comparability, ease of use, and practicality, paper-pencil tests of ability to apply facts and principles of human development were, therefore, constructed. This meant that data had to be presented and the student given an opportunity to make diagnoses and indicate a remedial procedure on the basis of those data. In order that the test would approximate a life situation, it was necessary to see that the facts were inter-related and presented in such a manner that they would be seen as clearly plausible and in the same perspective as a life situation. An artificially constructed case study with a pre-determined coverage made by the examiner and containing questions of remedial and diagnostic import seemed the most likely possibility. The scoring could be made objective and the coverage could be made to parallel the content usually included in courses and texts on adolescent development and behavior.

Measurement by the use of a case study test is not by any means a new idea. Home-made tests containing problems with alternative courses of action or diagnosis, and in some cases complete or partial case studies, were used in a limited way by college instructors who wished to improve their evaluation of students' learning. But, for the most part, the use of such case studies was based upon empirical grounds, and little, if any, research was done to assay their effectiveness or validity. A main criticism has been that students were asked to draw conclusions from too few or superficial data.

Baller (1) of the University of Nebraska, while a fellow with the Evaluation Service of the Commission on Teacher Education, pioneered in the publication of an authentic case study test (1) in three parts with both diagnostic and remedial questions after each part. The *Case of Mickey Murphy* was a definite and encouraging step forward in classroom evaluation. Baller has not, however, as yet published reports of any experi-

mental use of *Mickey Murphy* although he does cite some tentative norms.

### *Description of the Cases*

A case study, to be useful as a measure of ability to apply the facts of adolescent development, must be comprehensive in scope and in content. For example, if the case study were to deal primarily with emotional problems, success in reacting to it would indicate functional understanding of only one specialized aspect. One might reasonably ask whether the individual would show similar success in analyzing other basic problems. Consequently, to insure comprehensiveness, an analysis was made of selected course outlines and prominent textbooks in adolescent psychology, and mental hygiene for teachers. An analysis was made of the subject matter and the point of view expressed in common and the total relative amount of space given to each topic was noted. There appeared to be three major aspects of adolescent behavior and development: the physiological, psychological or emotional, and social. The three aspects are, of course, inter-related. In deciding on the coverage for each test each of the three aspects was taken as a major heading, and two questions were asked under each "What are the facts?" and "What are the problems?"

Further analysis of the material appearing under the three foregoing headings led to the compilation of points of reference with which to view adolescent growth and development, and they in turn were used as the criteria of coverage for the three tests to be constructed. Thus, the case study tests involve most of the principles and problems commonly presented in courses in adolescent psychology. At the same time, each case had to be plausible and realistic in order to provide a valid test situation. In writing such cases, a major problem is that of sharpening the implications of growth without making them stand out unduly.

To meet the requirements of coverage and plausibility the following three cases were constructed: the cases of *Barry Black*, *Connie Casey*, and *Sam Smith*. The major problem of *Barry Black* was social-emotional. While there were other precipitating conditions, the crux in the case centered in conflict,



frustration, and insecurity in social situations. The major problems of *Connie Casey* were caused mainly by physical and economic factors. The major problems of *Sam Smith* centered in intellectual-academic factors, but included a perifora of concomitants.

In addition to the main problems there were subordinate factors of age, intelligence, and school and home status. Barry Black, fifteen years old and in junior high school, was of normal intelligence; Connie Casey, eighteen years old and in senior high school, was of superior intelligence; Sam Smith, twelve years old and in elementary school, was of dull-normal intelligence. Barry's socio-economic status was average, Connie's below, and Sam's above average.

The first step in constructing the case studies was that of blocking out major subject-matter headings and problems to be included in each. The coverage outline of the criterion test was used as a guide. Excerpts having to do with the subject matter to be included in the cases were taken from over two hundred actual case studies.

When necessary, additions and amplifications were made to point up the cases. In order to avoid artificiality the cases were reviewed by a selected group of educational and clinical psychologists. The result was a comprehensive picture of the facts and problems of adolescent development from a wide selection of actual cases.

In dealing with an individual in a life situation, the psychologist or teacher is often compelled to make a diagnosis or to institute remedial procedures on the basis of a small amount of information. Later, changes may have to be made in the original diagnosis and remediation as new information becomes available. It was believed advantageous for the appearance of the actual situation and for future research to present each case in three sections. The student was asked to read and react to Section I before reading Section II, and to read and react to Section II before reading Section III. This would serve as an index of the effect of added information upon ability to make a diagnosis and to institute remedial procedures.

Three major sources of information were used in each case—

home, school, and community. They were presented in a different order in each of the three cases.

The next step was that of writing questions which would give the student an opportunity to display his understanding of adolescent development. Each section of each case study was followed by a group of questions having to do with diagnosis and a separate group of questions having to do with remediation. In analyzing the responses this would make possible a breakdown into three aspects: performance on the case study as a whole, ability to make a diagnosis, and ability to choose remedial procedures.

The diagnostic questions were set up in such a manner that the student would be given an opportunity to react to a hypothesis on the data given. He would also be given an opportunity to show his realization of the need for further information. The remedial questions would give the student an opportunity to make wise use of data in coming to decisions.

For objectivity, ease of scoring, and comparability of forms, objective type questions were used. However, with objective type questions, particularly those requiring a right-wrong answer, there is danger of superficiality. In life, in dealing with human beings, answers are seldom wholly right or wholly wrong—they are usually more or less right or wrong. The multivariate type answer seemed the best solution to the problem. Therefore, the student was given an opportunity to check *True*, *Possibly True*, *No Evidence*, *Possibly False*, and *False* on the diagnosis questions. A typical question is: "Connie is more than usually self-conscious and sensitive for a girl her age."

For the remedial section the student was given an opportunity to check *Strongly Agree*, *Agree with Reservations*, *Undecided*, *Disagree with Reservations*, and *Strongly Disagree* for some of the questions; and for others, *Very Advisable*, *Advisable*, *Undecided*, *Inadvisable*, and *Very Inadvisable*. A typical question in the remedial section is, "His remaining a Y.M.C.A. member should depend on attendance at non-gym activities."

Items were written which would be best calculated to bring out the various aspects of the three case studies so that the scheme of the original coverage would be preserved. The ques-

tions as written were presented to psychologists and selected graduate students for their reactions. Changes were made in accordance with their suggestions.

The cases and questions were mimeographed and administered to the writers' classes in adolescent development. Changes were made as a result of this class experience and the cases were re-mimeographed to incorporate improvements. These changes had mainly to do with the elimination of ambiguities, questions that were susceptible to misunderstanding, etc. The cases as finally mimeographed carried approximately a thirty-five per cent overload of questions so that later revisions might eliminate weaker questions without unduly interfering with the total picture of the tests. In their present printed form, the overload has been eliminated and the average student is enabled to react to each one in a little less than one hour. The case of *Connie Casey* has about 3000 words in the case data and from 25 to 30 diagnosis questions after each part, and from 12 to 20 remedial questions after each part.

### *Scoring and Keying the Cases*

For keying, the case studies were sent to a selected group of experts who were to read and react to them by marking the diagnostic and remedial questions in the same manner in which students were to answer them. A key was then made in accordance with the composite expert opinion. Each case study was sent to twelve experts. Eight answered *Barry Black* and ten each answered *Connie Casey* and *Sam Smith*.

An examination of the answer sheets turned in by the experts indicated some disagreement among them as to the correct answer. On the whole, however, agreement was fairly general, and the differences that did exist tended to be in degree rather than in outright conflict. It was interesting to note that experts whose training had been along "practical" remedial lines tended to agree among themselves, and psychologists whose training had followed along more theoretical lines tended to agree among themselves. On only about ten per cent of the questions was there really wide disagreement. The majority of such questions were later discarded in the process of refining the tests. There

were, however, few questions upon which the experts were in unanimous agreement, though a trend toward agreement or disagreement was to be noted. This situation was not unexpected. In a life situation, as was indicated earlier, there tend to be very few completely right or completely wrong answers. A great deal would depend upon the interpretation of various factors presented. It is a recognized fact that we tend to interpret in terms of our interests. For example, in the case of *Connie Casey* there are two different possibilities. First is the possibility that Connie is primarily a case of physical difficulty with accompanying lack of vitality. Second, is the possibility that Connie is primarily an emotional problem with socio-economic complications. Whichever major interpretation is accepted (both are in the picture) that interpretation will color both diagnosis and remedial procedure. It is interesting to note that public health nurses, on the whole, tended to interpret Connie as a case of physical difficulty; teachers tended to interpret her case along socio-economic lines; and the experts tended to stress the emotional instability aspect. In setting up a key, it was decided in all fairness that no expert would be discarded if he deviated consistently from the others.

It was also decided to set up a scoring system that would take into consideration expert agreement and disagreement. The assumption was that where one expert was in definite disagreement with the others there existed in the case evidence for that particular response.

There was also the aspect of class discussion. It was hoped that the case study instruments would have real instructional value in class. Insofar as that is true, it would be advantageous to have a scoring system which would allow the student to actually see himself in relation to the various experts. This would mean that the scoring system would, of necessity, be complex.

Balanced against the advantages of complex scoring was the disadvantage of an elaborate and time-consuming scoring system and the fact that tests of validity have in general failed to show any considerable benefit from differential weighting of items.

However, there is the possibility of a patent scoring key and the precedent of personality tests which have usually utilized a weighting system of one kind or another. The *Bernreuter Personality Inventory* (2) and the *Strong Vocational Interest Blank* (7) are cases in point. As Guilford (4) points out, "There are instances in which weighted scoring has materially improved reliability over that attainable with unweighted scoring."

Despite its disadvantages, the complex system was retained in view of its superior possibilities for class discussion and the possibility that it might give a more sensitive appraisal of the quality of the responses to test items. In the system finally selected, scoring was based on deviations determined by the magnitude of expert agreement or lack thereof. Two things were considered in arriving at a score value for each item—a modal deviation value and an expert selection value.

It may be recalled that each item is reacted to by checking one of five categories. In scoring the tests the following procedure was adopted. A distribution of the experts' marks on each item was made. Instead of using these actual numerical figures, the numbers of experts were grouped as follows. If 1 or 2 experts were in the same category a value of 1 was given to the category. If 3 or 4 experts were in the same category the category value became 2. If 5, 6, or 7 experts were in the same category, the category value became 3. If 8, 9, or 10 experts were in the same category, the category value became 4. The modal category (in which the largest number of experts appeared) was selected, and a value of 0 was assigned to the mode and a deviation score ( $-1, -2, -3, -4$ ) was assigned to every other category, according to its deviation from the mode. This was done in order to give partial credit to those answers which came closest to expert opinion, and to penalize those responses which deviated most from expert opinion. The modal category was assigned to a value of 0, and the deviations on either side values of  $-1, -2$ , etc., depending on how far they deviated from the mode. In the case of a bimodal distribution, each mode was assigned a value of 0 and the category on either side a value of  $-1$ , etc. The two score values (the one arrived at for the

number of experts marking it, and the deviation value from the modal category) were then added algebraically. Thus, the final score value for each of the five possible responses to the item was derived.

### *Reliability of the Case Study Tests*

The reliability of the three case study tests was initially increased by an internal consistency analysis and the discarding of items tending to lower reliability. Items that did not discriminate between those persons receiving high and those persons receiving low scores on each test and on the two parts of each test were discarded.

Each of the three tests was then administered to a population of 100 university juniors and seniors. The tests were scored and product moment coefficients of correlation were secured

TABLE 1  
*Split-Half Reliabilities of Connie Casey, Sam Smith, and Harry Black*

	Whole	Diagnosis	Remedial
Black .....	.79 $\pm$ .038	.68 $\pm$ .054	.64 $\pm$ .059
Smith .....	.73 $\pm$ .046	.72 $\pm$ .047	.55 $\pm$ .07
Casey .....	.77 $\pm$ .041	.71 $\pm$ .049	.64 $\pm$ .059

between the odd and even items for both the remedial and diagnostic sections of each test, and also for each of the three tests as a whole. The coefficients of correlation were corrected for length by the Spearman-Brown formula. The split-half reliabilities of the three case studies are given in Table 1.

The reliability coefficients of the three case tests are low in comparison with standardized tests of simpler function, but when compared with other tests dealing with complex human behavior, they are reasonably satisfactory.

### *Validity of the Case Study Tests*

The validity of the three case studies as measures of ability to apply facts and principles appeared to be established by their construction, coverage, expert scoring, item consistency, reliability, and utility.

Earlier in the discussion it was explained that the case

studies were constructed to cover the major problems and facts commonly presented in texts, courses, and discussions on adolescent development. This was done by making a careful analysis of such courses, and synthesizing, in outline form, their major areas of coverage and emphasis. The three case studies were then carefully constructed to parallel the outline, so that when completed they would contain in inter-related form, the coverage of the outline. This procedure contributed to "operational" (6) and "construction" validity. By building the tests to contain intrinsically accepted information and problems and to provide the student an opportunity to apply the content to remedial and diagnostic questions, operational validity was promoted.

That the answers given would be considered properly, a scoring system based on expert opinion was used which would tend to assure professionally "acceptable" interpretations.

The attempts at insuring item consistency have already been described under the section on reliability where it was explained that items that failed to discriminate between the more and less successful persons on the entire test and on the part tests, were eliminated. Hence, each item tends to be positively related to the over-all objectives of the test.

In general, the tests presented a realistic picture of an individual existing in a complex environment. The individual is confronted with precisely the same problems that an individual might meet in actual life, with all of the accompanying complications and ramifications, together with the impact of other personalities who in turn have their problems. The student is given an opportunity, as he would have in real life in a professional capacity, to interpret or diagnose the situation and then to recognize the goodness or badness of a wide range of remediation. The cases were carefully constructed to give a wide range of problems. The student is thus given an opportunity to apply what he knows to a life situation.

#### *Inter-Relationships of the Cases*

If it is to be assumed that each of the three cases is measuring ability to apply facts per se, then it would be expected that

a high positive correlation between the tests would exist. If, on the other hand, it is remembered that each case deals with a different aspect of adolescent development, the question might arise as to whether ability to apply knowledge about an emotional problem would indicate ability to apply knowledge about a social or a physical problem. There is also the question as to whether ability to apply knowledge about a school situation would indicate equal ability to apply knowledge about a home or community situation. If certain basic factors are involved, a positive, though not necessarily high, correlation might be expected.

As a matter of fact, with 67 cases the product moment coefficient of correlation between *Barry Black* and *Sam Smith* was  $.55 \pm .09$ . With 68 cases the correlation between *Barry Black* and *Sam Smith* was  $.39 \pm .10$ . The correlation between *Sam Smith* and *Connie Casey* was  $.62 \pm .09$ . These low relationships indicate the need for broad and thorough training of those who work with children, whether teacher or clinician.

### *Standardization*

'Tentative norms' have been prepared for the diagnostic and remedial sections of each of the case studies, and for each of the case-study tests as a whole. It is interesting to note that some students do very well in diagnostic and remedial items from Part I, but as the data become more complex and complete, they do less well. It is interesting to note that ability to diagnose accurately does not necessarily indicate good judgment in selecting appropriate remedial procedures.

### *Relationship between the Cases and Other Measures*

The *Ohio State University Psychological Test, Form 22*, and the *Case of Connie Casey* were administered to sixty-one highly selected college juniors. The correlation between intelligence and the score for the entire test was  $.23 \pm .122$ ; for the diagnostic section  $.10 \pm .127$ ; and for the remedial section  $.25 \pm .121$ . This, in the light of higher intercorrelations between case-study tests, would appear to indicate that the tests measure factors

<sup>1</sup> On four hundred to eight hundred cases for each test.



other than intelligence and that within limitations a more intelligent college student might or might not do as well as a less intelligent one.

A coefficient of correlation of  $.38 \pm .112$  was secured between the final marks of 59 students in a course in Adolescent Development and their scores upon the *Case of Connie Casey*. The case was not used in arriving at their final marks. Here again, factors are being measured by the case study which are apparently not taken into consideration by the instructors in grading student achievement. Troyer (8) has indicated a tendency for grades to be based on achievement of course goals that fall short of application of knowledge.

### Summary

This discussion explained the construction of three standardized case-study tests prepared to determine the extent to which students could use knowledge of human growth and development in making diagnoses and choosing appropriate remedial procedures. Refinements which led to increased reliability and validity were also described. There was comparatively little relationship between the cases and final marks in a course in Adolescent Development, or between the cases and increments of intelligence among college students. The relationship among the case studies themselves was positive but not particularly high, indicating that application of knowledge to each case tends to be highly unique.

The need for instruments which will measure the application of fact and principle was discussed. The instruments have been found useful in undergraduate and graduate courses in Adolescent, Educational, and Clinical Psychology and Mental Hygiene. They could also find use in in-service programs for the education of teachers, school psychologists, psycho-educational clinicians, and social case workers. The cases may be used both for instructional and evaluative purposes.

### REFERENCES

1. Baller, Warren. *The Case of Mickey Murphy*. Lincoln, Nebraska: University of Nebraska Press, 1943.
2. Bernreuter, Robert. *Manual for the Personality Inventory*. Palo Alto: Stanford University Press, 1935.

3. Eurich, Alvin C. and Cain, Leo F. "Prognosis." *Encyclopedia of Educational Research*, W. S. Monroe (ed.). New York: Macmillan, 1941.
4. Guilford, J. P. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill Book Company, 1942.
5. Horrocks, John E. "The Relationship between Knowledge of Adolescent Behavior and Ability to Use Such Knowledge." Accepted for publication in the *Journal of Applied Psychology*, 1946.
6. Jones, William J. "Improving Teaching in Ohio Through Experimental Use of a Cooperatively Developed Anecdotal Observation Record." Unpublished Doctoral Dissertation. Columbus: The Ohio State University, 1941.
7. Strong, Edward K. *Vocational Interests of Men and Women*. Palo Alto: Stanford University Press, 1943.
8. Troyer, Maurice E. "An Attempt to Improve the Comprehensive Examination at the Masters Level." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 235-248.

## SELECTING PERSONNEL WORKERS<sup>1</sup>

HARRY C STEINMETZ

San Diego State College

JOB analysts have remarked on the functional variability of secretaries—from the executive to the playful—so that a single set of man specifications is virtually impossible. The situation is the same among “personnel workers,” varying from executives and technicians to timekeepers and labor reporters. However, if patterns of such a complex thing as marital aptitude can be ascertained,<sup>2</sup> then some reliable differences between “good” and “bad” personnel workers may be discoverable.

Negligence in seeking these differences, in validating criteria of selection, is curious in view of the main job of the personnel research worker. In all other fields he seeks objectively significant ranges of abilities—observed output, tested accuracy, and rated teamwork—as criteria for agreement with which to select tests for employment, placement, promotion, and so on. It would seem reasonable to suggest that personnel workers and psychologists apply their lore to themselves.

In the absence of objective evidence for criteria, we have little ground for protest if jobs have gone to Tom, Dick, and Becky, and if training has been an unctuous elaboration of arm-chair reflection. Discussions of the qualities essential in personnel workers have been pitifully unconvincing. Without criteria of the worth of the worker, a discussion of his ideal qualities is reminiscent of the notorious ratings of salesmen from their appearance without regard to their sales records or of cigarettes without the blindfold test. It follows that disserta-

---

<sup>1</sup> Revision and abbreviation of a paper read on April 6, 1946, at the Regional Conference of the Council of Guidance and Personnel Associations at the University of California at Los Angeles.

<sup>2</sup> Adams, Clifford R. “The Prediction of Adjustment in Marriage” *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 185-193.

tions on the preparation of the counselor, the interviewer, the supervisor, and others, are largely abstract and presumptive.

Psychological understanding of all professional fitness will remain deficient, we dare say, until a study is made of those who are successful and unsuccessful in similar circumstances and under similar requirements. In the interest of stimulating empirical study, we propose four criteria of success in personnel work for trial in various circumstances and requirements. It is regrettable that industries and businesses and labor unions which maintain economics, physics, and other research departments nevertheless seldom sponsor genuine personnel research. It would seem reasonable to suggest that psychological and personnel associations should set an example. The four suggested criteria are these:

1. *A record of having adjusted others.*
2. *A good standing among associates.*
3. *A useful knowledge for the position.*
4. *An insight in terms of prediction.*

The first three will be discussed briefly and only insofar as they apply uniquely to the personnel worker.

1. *A record of having adjusted others.*—Experience is a criterion too superficially used in the evaluation of both the group and the individual personnel worker. How many men and women an employment manager has hired (perhaps with production on a cost-plus basis) is much less relevant than his comparative record of turnover and reshuffling. The duration of counseling practice is far less important than the evaluation of success. We know nothing of the significance of self-adjustment among personnel psychologists. The experienced personnel executive should have a record, and be entitled to a record, of having adjusted others.

2. *A good standing among associates.*—Present standing among associates requires a three-way rating. Personnel workers no more than teachers can be reliably and validly evaluated on the basis of supervisory ratings alone; associates and subordinates have as significant viewpoints if not relations and they are more numerous. This is the very basis of Anglo-

Saxon justice that entitles a man to his day in court under judgment of his peers.

Distributive business long ago learned that no realistic appraisal can exclude the opinions of the consumers. Counselors and supervisors and foremen who rate employees should be rated by them, if only so that top management may see how management appears to employees at the level of immediate contact. In a large organization, particularly in a large office force, there can be no more convincing criterion of general worth than becomes available as a result of mutual rating.<sup>3</sup>

A category which could plausibly be included in three-dimensional ratings of and by and for personnel workers is this: "To what extent is Mr. John Doe concerned with the development of others toward full and efficient use of their abilities?" Surely the personnel worker must satisfy those for whom he works, with whom he works, and on whom he works. The extent of unanimity of judgment may indeed prove a cardinal test of an organization.<sup>4</sup>

3. The criterion of useful knowledge would seem to call for an examination prepared by "outside" experts employed to job analyze a specific personnel work situation. There is both general and local information which it would appear valuable for interviewers, supervisors, and rating personnel to have. This has been well proved by Quentin File in "Measurement of Supervisory Quality in Industry" which resulted in the only standardized test in this field and one which is quite relevant for personnel workers.<sup>5</sup>

4. A perhaps novel but, we believe, promising criterion for the evaluation of personnel workers derives from the measurement of their insight into others as evidenced by their prediction of responses. This we have called "Measuring Psychological Understanding."<sup>6</sup>

<sup>3</sup> Henry Wood Shelton, 1560 Torrey Pines Road, La Jolla, California, has prepared mimeographed material on Mutual Rating. He is a retired efficiency engineer (Taylor system).

The author of this article has also written a *Manual of Industrial Efficiency Rating* which may be procured from him at 3518 Union Street, San Diego 1, California.

<sup>4</sup> Unanimity (reliability) validates the usual validating criterion, ratings. When tests are validated and weighted they may in turn be used to test accuracy in the placement of personnel workers.

<sup>5</sup> *Journal of Applied Psychology*, XXIX (1945), 323-337.

<sup>6</sup> *Journal of Clinical Psychology*, I (1945), 331-335.

Adequately reliable estimates of understanding of psychological categories (*verstehen psychologie?*) may be derived from the relative ability to score high or low in a personality test following special instructions. What is sought is the ability to predict the scoring key.<sup>7</sup> We are engaged currently in an analysis of the ability to evince insight through this unorthodox use of the Guilford Martin *Inventory of Factors GMIN* which measures general activity, ascendancy, sex preferences, inferiority feelings, and nervousness.<sup>8</sup> The test is being administered naively, followed by ratings of self and others and of knowledge of categories AMN<sup>9</sup> (the epistemological elements), and then readministered under instructions to answer as persons X and Y have answered, and finally readministered under instructions to appear maximally dominant-masculine-composed one time and submissive-feminine-nervous another time.

The research just mentioned indicates the use of unorthodox testing for sampling insight into persons as well as the understanding of categories. Two such procedures from private practice will be reported.

... in outlining a selection technique for a small manufacturer, the suggestion was made that he use his own responses in the *Strong Vocational Interest Test* as a scoring key. It is believed that the information secured from such a system of matching may be more revealing than comparison of total raw, percentile or other scores. In any event, the manufacturer secured a sales manager (with whom he had to work closely) who not only satisfied arbitrary standards of knowledge and personality which had been agreed upon in advance, but who also was extremely similar to his employer in interests, tastes, preferences, and attitudes. One can think of many other situations wherein a similar procedure might prove to be advantageous.

However, more important than true similarity may be ability to duplicate the other's responses and so demonstrate insight. This may be on an individual or group basis. After

<sup>7</sup> Steinmetz, H. C. "Measuring Ability to Fake Occupational Interests." *Journal of Applied Psychology* XVI (1932), 123-130.

Kelly E. L., Miles, C. C. and Terman, L. M. "Ability to Influence One's Score on a Paper-and-pencil Test of Personality." *Character and Personality*, IV (1936), 206-215.

<sup>8</sup> Sheridan Supply Company, Beverly Hills, California. We are indebted to Dr. J. P. Guilford for tests used in this experiment.

<sup>9</sup> Factors *G*<sub>1</sub> and *I*, less reliable and communicable, are not being used but their influence and the influence of "empathy" upon them will be noted.

the vocational interest test had been given the manufacturer's applicants (supposedly naively, although there may have been some fudging in the direction of aptitude-for-salesman), the manufacturer himself might have appeared before the group of assembled applicants and delivered a short talk, following which the applicants might have been asked to approximate his responses to both interest and personality questions. Psychological perception of the other person is certainly a factor in getting along with him, given the motive to do so; this is the crux of the matter.<sup>10</sup>

If we were a chief probation officer, we should be less interested in a deputy's book learning<sup>11</sup> than in his insight into the minds of delinquents as revealed by his ability to duplicate the average or modal responses of youths in temperament, character, and interest, and opinion batteries.

In marital counseling, after talking with each mate, we sometimes write significant questions on 3 × 5 cards. After each has sorted them into "yes" and "no" piles, we ruffle his "yes" pile and make a blue line down a reverse edge, and then ruffle her "yes" pile and make a red line down the same reverse edge. Then we shuffle the cards and have each re-sort as he thinks the other has just answered. On the reverse of another edge we mark his "yes" cards red and hers blue. We now have the opinion of each, the comparative knowledge each has of the other, and by implication, at least, the opinion each has of the other. We are not yet ready to generalize, but believe that we are gathering evidence that where there is a twenty per cent or greater difference in knowledge of one by the other, the main problem seven or eight times in ten lies with the one who least well knows the other; certainly the counseling of this one, for insight, proves the most difficult, but this evidence of need opens the way.

This last criterion comes to grips with the need for performance tests in personnel psychology, the need for ways and means of selecting and training the selectors and trainers. Its underlying assumption is that the proof of knowledge here, no less

---

<sup>10</sup> Steinmetz, H. C., *op. cit.*

<sup>11</sup> Gillin, J. L. and Hill, R. L. "Success and Failure of Adult Probationers in Wisconsin." *Journal of Criminal Law and Criminology*, XXX (1939-40), 807-829, reports the education of supervisors as eleventh among twenty-eight factors influencing probation.

than in meteorology and medicine, is prediction. A diagnosis that cannot be tested is so much conversation. There is an element of prognosis in every diagnosis and only so does descriptive verbiage in psychology have utility.

The flight of psychiatrists, psychologists, and personnel people from prediction tests of their descriptions of individuals and groups is nothing short of scandalous in our opinion. Offices are full of diagnoses (insights?) which are never checked by follow-up; this despite the fact that every interviewer's recommendations are virtual predictions.

We are reminded of the clinical psychologist who said that after 30 years of counseling he had given up marital and premarital work, for all of the marriages he had advised had ended in divorces, as had his own and his children's, and all whom he had sought to obstruct had been consummated happily, while divorces he had sanctioned had ended in remarriage of the same couples. One can only wish that this psychologist had checked on himself, or been checked by some agency, and quit much earlier. It will not do for a non-directive or yes-yes counselor to smirk at this story, for he or she should terminate every interview with a prediction for later check-up. The judgment of personnel workers should be nicely revealed from a file of predictions, even if guardedly conditional, which may be periodically reviewed. There is a statistical significance to interpersonal judgments which cannot be neglected, as Theodore R. Sarbin rigorously proved in "The Logic of Prediction in Psychology."<sup>12</sup>

Nearly all examinations in psychology and personnel work measure only knowledge of categories (a la mode) and of practices (mainly current). Other criteria for recognition and responsibility are time service and publication. None of these is an adequate substitute in artisanship for a more tangible evidence of competence. We have emphasized the measurement of insight for the sake of stimulating attention to the need in psychology and personnel work, for the sake of the public and our professions, of performance tests of ability. Psychological insight is practically meaningful only insofar as it is evinced in

<sup>12</sup> *Psychological Review*, LI (1944), 210-227.



tests of ability to duplicate or predict the responses of individuals or groups, which last includes the duplication of valid scoring keys.

In this essay we have commented on the incongruous neglect of objective evidences of personnel abilities; four criteria are proposed, of which three common ones are redefined; and a fourth criterion (or test bearing face validity), insight evidenced by prediction of behavior, is reported to be measurable, useful, and worthy of analysis.



# WHAT CONSTITUTES A MINIMAL SCHOOL TESTING PROGRAM

WALTER N. DUROST  
World Book Company

## *Minimal for What Purpose—for Whom?*

BEFORE one even can begin any intelligent discussion of the minimal testing program, it is necessary to define the terms involved. First it must be understood that we are talking about standardized testing and not the overall evaluation of the outcomes of education. The use of standardized tests constitutes only one part of the total evaluative process. By a minimal testing program we mean the least amount of *standardized* testing which is consistent with good school administration in the light of all local circumstances.

For any testing whatsoever to be justified in any school situation, that testing must be purposive. It is a waste of time to administer tests and to score them and then to file the results away in some dark corner to collect dust. Before a test is selected, the purpose for which it is to be given should be clearly in mind, and after it has been given, it must be followed up to see that the purpose has been achieved!

Standardized testing may have many purposes. The administrator may wish to get an overview of his administrative unit from the point of view of the general level of ability or of the achievement in any one or several subject-matter areas. For that purpose, he selects appropriate tests and sees that they are administered in accordance with the proper standards, and the results are analyzed to provide him with the information he needs for determining grouping and promotional policies.

The supervisor may wish to have the advantage of standardized test results in determining those areas which require emphasis in her supervision, or in determining those schools

where her assistance is greatly needed, or in determining in which of several areas steps should be taken to initiate curriculum revision.

The teacher may wish to initiate testing because she needs to know at what rate her pupils can be expected to absorb her instruction and at what level this instruction should be pitched. She may wish to know basic information which she can count on as having been adequately taught at earlier grade levels. Or she may need the benefits of testing to help her organize her class into sections for more effective instruction.

In many cases testing is multi-purposive in that it may serve the administrator, the supervisor, and the teacher, and that program can be considered most successful in which every person in the administrative setup has gained the maximum benefit from its application.

However, in this article the focus of attention is not the administrator, nor the supervisor, nor even the teacher, but the pupil. What constitutes a minimal testing program from the point of view of the pupil? How much testing is needed in a typical school environment in order that a pupil may be fairly and honestly dealt with, given the maximum opportunity to obtain an education or, more broadly speaking, to develop and apply his natural talents?

### *The Cumulative Record*

A pupil-centered testing program is of little value unless the local administration has made effective plans for the recording of test data on a cumulative basis. Because of the unreliability inherent in any test score, the single test score always must be interpreted with caution. Two IQ's recorded on a cumulative record card are worth much more than two separate measures interpreted without reference to each other. This is just as true of achievement measures or of interest and personality test results.

### *Intelligence Testing as a Part of the Minimal Program*

Intelligence tests constitute one important part of any minimal pupil-centered testing program. To support this statement

it is necessary for us to consider briefly what an intelligence test really is. Many definitions have been written, but for our purpose perhaps the following simple statement will suffice. An intelligence test is a carefully selected series of problem situations for the solution of which varying amounts of mental ability or skill are required; the score on such a test then must be interpreted by referring the pupil's measured ability to some standard of comparison or norm. A study of this definition will make evident several things that intelligence tests definitely are not. An intelligence test is not a measure of inherited ability independent of environmental factors. An intelligence test is not a measure entirely free of the effect of school attendance or of other formal instruction. An intelligence test is not a measure of mental ability independent of motivation. Therefore, an intelligence test is invalidated to the extent that the environment is atypical, to the extent that the school experience has been other than normal, and to the extent that the emotional attitude of the child at the time the test is given is not favorable.

As defined, the intelligence test is comprised of a series of problem situations involving the exercise of mental ability in their solution. However, life is so full of problem situations that the variety of such that might be presented to a person is almost endless. In a practical situation, the type of problem occurring in intelligence tests has become more or less fixed by both the pattern of the past and by the application of logical criteria in the determination of test content. In recent years, the application of highly involved and technical statistical analysis has been applied to the problem of the appropriate content of intelligence tests with the result that in some tests the so-called primary mental abilities are now measured by separate tests. However, even these tests do not vary greatly from the pattern that has been established over the past twenty-five years, and they differ chiefly in that each ability is recognized as a more or less independent mental attribute, and thus is given the dignity of independent measurement and interpretation. Assuming, of course, that the mental abilities so measured are the ones which are most important for school and life

success, this trend toward the separate measurement of more or less independent abilities is a praiseworthy one, and one which it is hoped will become more widely accepted in the future. Its wider acceptance is conditioned by the willingness of test users to devote the necessary time to such a program.

### *Varieties of Intelligence Tests*

Current intelligence tests differ enough one from another so that it is not practical or reasonable to talk about *the* IQ, but one must always refer to an IQ derived on a specific test. Thus, one speaks of an Otis IQ, or a Terman-McNemar IQ, or a Kuhlman-Anderson IQ, or a Stanford-Binet IQ, or a Pintner Non-Language IQ, or a California Mental Maturity Test IQ, fully recognizing that without this further definition one cannot truly comprehend the significance of the intelligence measure derived. Let us note just a few of the sources of variation in IQ from different tests.

First, and most evident, is the variety of problem situations found in the test. For example, the Stanford-Binet includes a great many performance items in contrast to a general group test of intelligence which is entirely a paper and pencil situation. A non-language test of mental ability, such as the *Pintner Non-Language Test of General Ability*, is free of the effects of language entirely while a general verbal intelligence test, such as the Henmon-Nelson, samples a variety of verbal skills. Some group intelligence tests are of the self-administering type, while others maintain the pattern of separate subtests yielding separate measures which are combined into some kind of average. Some tests emphasize the speed factor, while others have generous time limits. The Binet assigns a mental age value to each new problem situation. Most group tests, on the other hand, translate only the total score into mental age or even skip this step entirely and derive a measure of relative brightness. The norm populations also vary greatly, in size, in geographic distribution, and in racial composition. And so it goes. These illustrations certainly should suffice to illustrate beyond question the importance of designating the test upon which every IQ is based.

*Choosing the Intelligence Test*

If intelligence tests differ so greatly in their content, how is one to know which test to use for any specific situation? Generally speaking, for use in schools the group test of verbal mental ability is the most useful, and is the one which is considered basic to our minimal pupil-centered testing program. In a sense, it can be considered a scholastic aptitude test rather than a measure of the ability of the individual to apply his intelligence to the solution of *all* types of life situations. In contrast, the individual test, such as the Stanford-Binet, measures the reaction of the child to a wider variety of problem situations, many of which are met outside of the purely verbal environment. The individual test has the added advantage of constant motivation, in that the examiner does not pass from one problem situation to the next until he has elicited from the pupil being tested the maximum performance which can be expected of him under conditions of good motivation. The non-language test of mental ability is a supplement to the verbal test, having great value as an indicator of those pupils whose language abilities have not been developed to the point where they can make a fair showing on a verbal test of intelligence. Cases of specific disability in reading or of children who are hard of hearing and, therefore, whose verbal environment has been limited, are illustrations. Thus the non-language test or the individual Binet may be considered for the purposes of our minimal pupil-centered program as supplementary tests to be used only as indicated, and not as instruments to be applied generally to all pupils.

*When Should Intelligence Tests Be Given?*

In determining the frequency of the administration of intelligence tests in a minimal pupil-centered program, a variety of factors must be considered. First, there is the unreliability of the test instrument itself. Any group or individual intelligence test is subject to error, not in the sense of a mistake in the administration or interpretation of the results, but in the sense of a variation in results from one testing to another due to the operation of chance factors. By the operation of these chance

factors alone, the intelligence test quotient may vary as many as ten or fifteen points in a substantial number of instances, where two or more tests are given. Thus, intelligence tests should be administered with reasonable frequency in order that a child may not be unduly handicapped, either by being overrated or underrated on a specific test. Furthermore, in every case where the teacher's judgment of the child's ability varies widely from the results of the test, a retest should be given immediately. Recognizing that in many situations administrative considerations make it impractical to administer intelligence tests annually (a recommendation for any *optimum* testing program), it is recommended that intelligence tests be administered at least four times during the child's school career. Grades in which these tests may be given with good results are Grades 1, 3, 6, and 9. These grades are chosen on the assumption that the tests are to be administered at the end of the school year for administrative reasons which will be noted subsequently. If the tests are given at the beginning of the year, the recommended grades might be 2, 4, 7, and 10.

#### *Advantages of Objective Intelligence Testing*

The statement is often made, especially by the older teachers to whom standardized testing of any kind is anathema, that they do not need an intelligence test in order to know who are the bright pupils in their classes, or who are the stupid pupils, but that their day by day observation based upon long years of experience gives them this information without benefit of any testing. Over the years the correlation between intelligence test results and teachers' judgments of intelligence has been in the neighborhood of .50. Thus, either the teachers or the tests are in error substantially, or both are in error to some extent. The validity of intelligence tests as measures of pupil brightness has been the subject of many researches, and this is equally true of teachers' judgments of pupils' abilities. Among the most noteworthy studies revealing the validity of the objective tests as measures of individual brightness are the researches of Dr. Terman, who made extensive use of test results in his genetic studies of genius. Dr. Terman found that he could rely



better on the simple technique of asking the teacher to list the name of the youngest pupil in her class than he could upon her judgment as to which pupil in the class was the brightest. If asked to name the brightest child too often the teacher would refer to him one of the older pupils whose life experience gave him greater absolute mental power, but not greater in proportion to his *age*.

Thus, we can say that teachers generally can identify the more able pupils in their classes, but that they tend to do this without regard to the factor of chronological age, basing their estimates upon absolute mental ability, instead of relative mental ability. To put the matter in testing terms, they are basing their judgments upon *mental age* rather than upon IQ. The objective standardized intelligence test yielding an index of relative brightness, generally the IQ, supplements the teacher's judgment and reveals those instances where unexpected ability is present but is obscured by the factor of lower chronological age, or where the learning potentiality is overestimated because the factor of overageness has not been taken adequately into account.

The objective standardized test of intelligence has the added great advantage that it offers much more nearly comparable opportunities for all pupils to reveal their mental ability since it is relatively free of the effects of specific teaching in school, and is not influenced by personal factors which might bias the judgment of the teacher. Furthermore it yields an index in numerical terms, making it possible to relate it easily to other variables.

#### *Limitations of the Group Test of Mental Ability*

General intelligence tests or tests of mental ability have certain definite limitations. For practical reasons the length of the test usually is limited to something under one hour and, therefore, the sample of pupil behavior is not measured with anywhere near perfect reliability. Also, because of the shortness of the test, the variety of mental abilities that can be sampled is limited. And finally, the general group test of mental ability, as well as all the other tests of intelligence, is affected

by environmental influences to a greater or lesser extent, depending upon the components of mental ability which the test presumes to measure.

### *How to Use the Intelligence Test Results*

It has been stated that the minimal pupil-centered testing program will call for the administration of a group test of mental ability at least four times during the school career of the child. It also has been stated that any testing is valueless unless it is purposive. How then, is the teacher, the supervisor, and the administrator to use the intelligence test results of a pupil-centered program so that they will be of greatest possible benefit to the pupil? This problem has many ramifications, but they can be summarized by saying that the intelligence test results should be treated as a measure of pupil potentiality in *every situation in which a decision has to be made concerning the school career of the child.*

Thus, the administrator considers the IQ of the pupils, either individually or collectively, in determining his policies of promotion and of grouping. The supervisor considers the IQ's of the children in the group under her control in determining the general level of instruction and the rapidity with which new instructional material will be introduced in the classroom. The teacher considers the measure of pupil potentiality in determining the level at which she will pitch her day by day instruction, and in determining the amount of differentiation that she must make in presenting her material to her class to take care of the needs of atypical pupils. Thus, to the extent necessary, she will individualize her instruction so that those pupils whose learning rate is slowest, as evidenced by the intelligence test, will have most opportunity to benefit by repeated instruction, while those whose learning rate is highest will be allowed to proceed at a rate consistent with their ability to learn. With each new intelligence test record entered upon the cumulative record folder, the knowledge of the administrator, the supervisor, and the teacher about the child *as an individual* increases and becomes more stabilized. As the pupil reaches the higher grades in those instances where the cumulative record has been

kept conscientiously and accurately, the measure of intelligence based upon at least four applications of group tests in controlled situations will allow for little variation and may be used as a basis for prediction of future success in academic work or in vocational fields, *provided* the limitations of these measures are fully realized. After all, success in any field depends only in part upon the *ability* of the person involved, and is conditioned further by the amount of effort extended and the degree of motivation present.

### *Minimal Achievement Testing*

Intelligence tests may be considered a measure of potentiality rather than achievement, and their primary use in the school situation is as school aptitude tests, so-called. The achievement test, on the other hand, measures the outcomes of instruction in the school situation. However, the standardized achievement test, such as the *Stanford Achievement Test* or the *Unit Attainment Scales*, is a measure of achievement generalized to fit the country as a whole, and is not a measure of specific instruction in any one community or state. It is important to realize this because otherwise one cannot understand clearly the function of the general achievement test in the local testing program. Perhaps one way to clarify this matter is to consider briefly how an achievement test is made and how it would differ from a locally made test covering the specific content of instruction.

### *A Concise Definition of a Standardized Test*

An achievement test may be defined concisely as a series of test questions rigorously selected to measure a typical cross-section of instruction in each subject-matter area and subsequently administered to large representative groups to provide a basis for the interpretation of test scores. For example, when you have an arithmetic score on a pupil it becomes meaningful only if it is based on an appropriate test for the grade and only when compared with the norm. Of course the same is true of all the other subtests. It is these techniques of curriculum analysis, item tryout, and administration to establish norms

which give the standardized achievement tests their unique place in the educational scheme. If the intelligence test and the achievement test used are both standardized on the same general type of population so that the values derived are more nearly comparable, the value of both types of test is greatly enhanced by this comparability.

### *Overview of Achievement Test Construction*

After the general content of the achievement test has been decided upon, the first step in constructing the test is to establish the ground rules, or criteria, on the basis of which items will be constructed for preliminary tryout. These ground rules will vary somewhat from test to test. For example, in a reading test, the authors must determine what particular aspects of reading are to be measured, and what type of test item is most suitable for the purpose in mind. For a subject such as arithmetic, the ground rules will emphasize the study of curricula and textbooks to determine the nature of instruction current in the typical community in any part of the country. A similar analysis of courses of study and textbooks is made for all the informational subjects and it is only after such extensive study that the test items are constructed.

The second step is the actual writing of test items and their editing, which involves the review of the items by a large number of people from different points of view. After the items have been carefully edited and are in substantially final form, an experimental edition is prepared and tried out as a basis for further statistical analysis.

The statistical treatment of the test data constitutes the third major step in the construction of an achievement test. The per cent of pupils answering each question correctly is studied, and items failing to show a reasonable gain from grade to grade are eliminated. Items which are too easy or too hard are taken out of the test and the forms are balanced so that each form is of approximately the same difficulty as each other form.

The fourth major step in the preparation of an achievement test involves the establishment of norms. To do this, a second

experimental edition is prepared of one of the forms, and these tests are given to extensive groups of children all over the country. On the basis of the data growing out of this standardization program, grade equivalents, age equivalents, percentile norms, and standard scores are all established. The final step in the preparation of the achievement test involves the writing of the final *Manual of Directions*, or *Directions for Administering*, which explains the administration and interpretation of the test, and the final preparation of Keys, Class Records, and other accessory material.

*What Achievement Tests Should be Included  
in the Minimal Program*

Any minimal program of pupil-centered testing should include achievement tests in the basic skills. Fundamentally, the schools still find their major responsibility in the teaching of reading, arithmetic, spelling, and in the mastery of the English language. Tests in the information areas and tests of other subsidiary skills are important too, but are definitely secondary to the basic skills mentioned above.

*How Often Should Achievement Tests Be Given?*

Ideally, achievement tests should be given annually, but in a minimal program where this is not possible, a standardized achievement battery should be given at least each time the curriculum changes notably. Thus, when the child finishes the primary grades, a standardized achievement test would be of great value in determining his relative standing in the various major subject-matter areas, both from the point of view of the administrative decision concerning promotion, and from the point of view of the direction the subsequent instruction should take. If the schools operate on a junior-high-school plan, a test at the end of the elementary grades again gives the child a fresh start when he moves into the new organization. When the child moves from junior high school into senior high school, a standardized achievement test should prove of great value in determining the nature of the course of study he should take and, generally, in planning his future school activities.

*Choosing the Achievement Test*

It may be helpful to enumerate some of the factors which should be considered in choosing an achievement test. These factors may be considered under two headings, factors bearing on the content of the test, and factors having to do with the administration of the test.

Under content, some of the following considerations should be noted.

1. Were the items selected on the basis of a careful examination of the curriculum in each subject-matter area?
2. Were the items retained in the final test refined by try-out and subsequent statistical treatment?
3. Are several forms available and are these forms equal in difficulty?
4. Are various types of norms provided and are these norms based upon adequate and representative population?
5. Does the publisher provide adequate manuals, keys, class records, etc., for utilizing the test results?
6. Does the test suit reasonably well the content of instruction in the local situation, making it practical to compare the results of local instruction with the norms provided?

Under administrative factors some of the following items should be considered.

1. Are the directions for administering the test clear and unambiguous?
2. Is the scoring of the test objective and are mechanical aids provided for making this scoring as easy as possible?
3. Are the time limits reasonable in terms of the amount of time available locally for the testing?
4. Are the claims for what the test may accomplish conservative and reasonable rather than exaggerated? For example, the claim is sometimes made that an achievement test is diagnostic because sub-scores can be derived on a wide variety of topics or sub-topics. When these claims are investigated it is found that the sub-topic scores are based on two or three items and cannot be depended upon.

*Who Should Be Responsible for Evaluating Tests  
and Planning Testing Programs?*

Only a moment's thought is necessary to convince anyone that the problems of choosing tests, of planning the details of test administration and analysis, and of evaluating for future reference new tests which may appear on the market are ones involving considerable technical knowledge and skill. Obviously it is not practical for every administrative unit to have a person spending full time on activities centered around the testing program. In most situations it would be possible for some one teacher to be designated to study the measurement field as an area of specialization and to devote some time to the testing activities of the administrative unit.

Such a person would be encouraged to study tests and measurements in summer school or graduate school with an added course or two in elementary statistics. In the organization of this person's work provision should be made for time to be spent systematically on testing problems with, possibly, 100% relief during periods when unit-wide programs are under way.

One of the responsibilities of such a person would be in-service training of teachers in testing methods and utilization. Not every teacher should or can be a test specialist but every one should know enough to know that she is hurting only herself, for example, if she coaches her children on a standardized test, or if she deviates from the Directions for Administering. Teachers may need considerable help in planning and carrying out their part of the scoring and analysis of test results, in making class item analyses, etc. All of this could be part of the responsibility of the test specialist.

*Mechanics of the Testing Program*

Any testing program, whether it be intelligence testing or achievement testing, will be successful and will be well received by the local teachers to the extent to which it is carefully planned and each person involved knows exactly what is expected of him. It is most helpful to make use of large manila envelopes as test containers for the distribution of tests to

each classroom teacher. By keeping the tests in these envelopes at all times they are handled without confusion and each step in the process from the administration of the test to its final interpretation can be controlled. Ordinarily it is a good idea to prepare a label which may be pasted to the outside of the envelope containing the identifying information for the class, school, grade, etc., and listing each step in the process from the time the tests are administered until they are finally interpreted and the results entered in the cumulative record for each child. Supplementary directions to teachers may be included in this envelope along with the standard directions coming with the test.

Clear, concise, simple but complete directions to the teachers involved in the administration of tests can do much to improve the quality of a testing program. In planning any testing program, the matter of motivation must be considered. If teachers feel that they are being imposed upon by being asked to give the tests and to score them, the results cannot help but be affected thereby. On the other hand, if teachers understand the positive gains to themselves in the "quality control" of their instruction, and are assured that there are no penalties involved, their cooperation may be secured to the maximum possible extent.

#### *Supplementary Testing in the Minimal Pupil-Centered Testing Program*

So far our discussion has centered around the use of intelligence tests and achievement tests exclusively in the *minimal* pupil-centered testing program. Are these the only tests that need to be included? For the vast majority of pupils the testing outlined will be sufficient, but for a minority, perhaps 10 to 15%, additional testing will be required from time to time. Under certain circumstances repeat testing with an alternative form should be done. For example, if the intelligence test results seem out of line with the teacher's judgment, a second form should be administered. Similarly, if the achievement test results do not check with the teacher's judgment as to what may be expected of the child, it is reasonable to follow this up by administering a second form. If the local organi-



zation provides for a specialist in tests and measurement, there may be occasions when personality tests, such as the *Washburne Social Adjustment Inventory*, can be given to great advantage, perhaps in connection with an educational clinic. Such tests should be used with caution and interpreted only by those who are willing to take the time to study carefully just what the test scores mean. In certain cases, longer and more diagnostic tests in the subject-matter areas are called for when special disabilities are suspected, and these properly may be considered part of the minimal pupil-centered testing program.

### *The Optimum Program*

So far we have concerned ourselves exclusively with the minimal program on the assumption that the local community has not progressed in its use of tests to the point where it may do as much testing as might be called for under an optimum program. An optimum program is one where tests are used wherever and whenever they are needed to obtain objective information concerning the pupils in the system. Under an optimum program annual achievement testing in all grades might be the practice rather than the exception. The optimum testing program might include the administration of teacher-made tests on a unit-wide basis. Such a program might include the routine administration of one non-language intelligence test during the course of the child's school period, possibly at or near the ninth-grade level. The optimum testing program might include the administration of special types of tests, such as tests of critical thinking at certain grade levels where the ability to think logically is being emphasized. In short, the optimum testing program would be the minimal testing program supplemented to include the administration of tests at other times and at other grade levels where the purposes to be achieved by the testing were clearly understood, and were considered to be of sufficient value to justify the expenditure of time and money.

### *Conclusion*

A testing program, minimal or otherwise, is a means to an end and not an end in itself. A fine scalpel in the hands of a

poor surgeon will not insure a safe or competent operation. Neither will a good test insure good results unless handled expertly and conscientiously. It has been said many times that teachers cannot be trusted to give and to use test results. That is a libel of the most flagrant sort on the intelligence, integrity, and honesty of those to whom we trust the well-being of our sons and daughters. It is equally unreasonable, however, for school administrators to expect teachers to train themselves in the proper attitudes and skills needed in the testing field. We do not leave curriculum content and methods to the individual teacher solely, even while encouraging initiative and resourcefulness in our instructors. Neither should the administration and interpretation of tests be left to chance. A crying need is for more in-service and out-service training of teachers and supervisors and administrators in this vital field.

•

# THE CRITICAL INTERPRETATION OF TEST RESULTS IN A SCHOOL SYSTEM

RAY H. SIMPSON  
University of Alabama

THE skill of many teachers and administrators has been woefully weak in properly interpreting standardized test results. In most cases this has not been due to a lack of potential ability on the part of the educators but rather to a lack of appropriate pre-service and in-service training in an understanding of the possible values and limitations of various types of tests. Largely because of this lack of understanding, some educators have attributed too much importance to particular test results while others, disillusioned, have said in effect: Away with all standardized tests.

Recently the writer came in contact with a situation which emphasized the point which has just been made. The facts in the case are essentially as given below, disguised only enough so that the school system itself cannot be readily identified. Following is a brief presentation of the situation together with questions and comments which are designed to be thought-provoking.

Last spring a large school system gave a comprehensive standardized achievement test to all of its seventh-grade pupils and found that the median score of those taking the test was 9.2 or 1.5 grades above the median score of 7.7 of the groups on which the test was standardized. Self-satisfaction verging on smugness was the typical reaction of the teachers and administrators in the system.

It is quite possible that a very superior type of work is being done in this school system, *but* before this could be known with any great degree of certainty, we would need some additional information. The questions and comments given below sug-

gest some of the areas which the thoughtful teacher and administrator should investigate and consider before drawing his conclusions with respect to the effectiveness of the school system.

*How much retardation is there in the school?* It is quite possible for a school system that is doing a relatively poor job to reach or even to exceed the test results of this school by having a rigid policy of promotion. For example, the writer is familiar with a school system which takes great pride in its "standards." With very few exceptions, no learner is promoted until he has reached the "norm" score for his grade on the standardized achievement test which is used as a primary basis for promotion. This means that a child in the fifth grade is not promoted to the sixth grade until he can make a score equivalent to 5.9 on the test. This policy has resulted in excessive retardation with about 50% of the seventh-grade pupils retarded at least one year. Obviously, a comparison of the median score of this group with the group on which the norms of the test were based is unsound. With excessive retardation even an efficient school system may seem to have a good seventh grade if only grade norms are considered.

*How much elimination is there in the school system before the seventh grade is reached?* Some schools have a deliberate policy of "weeding out" the weakest members of a particular age group; other schools just fall into the habit. Consider each of two schools which started out with 200 pupils entering the first grade. In school A only 60%, or 120 pupils, ever reached the seventh grade—the others were eliminated in one way or another. In school B 95%, or 190 pupils, entered the seventh grade. Since it usually is the weaker learners who are eliminated in a school such as A, it is obvious that school B with a lower median score than A might still be doing a much better job of educating the children in its community. The holding power of a school needs to be carefully considered in interpreting test results. The twin evils of excessive elimination and excessive retardation demand thoughtful consideration and test makers and publishers should be encouraged to give figures indicating the extent of the elimination and retardation in the schools on which the norms are based.

*Did the teachers teach the test directly or indirectly?* It may seem undignified even to suggest that such an unprofessional practice might be carried on. But in certain school systems the practice is carried on by certain teachers, and those who are trying to interpret test results need to be aware of this possibility. Of course any time a group of learners is taught a test the use of norms accompanying the test becomes meaningless. Unfortunately some administrators and supervisors have unwittingly encouraged this practice, partially through a procedure suggested by the next question.

*Are teachers given raises or promotions on the basis of test results?* Some superintendents, principals, and supervisors casting about for an objective basis for giving promotions in rank or salary increases have settled on the idea of giving these rewards to those who can produce the best test results. The goal is admirable but this particular method has resulted in many unprofessional practices and should be eliminated in any place it exists.

*Is the primary goal of the teachers to help the learners or to get good test results?* In some school systems it is unfortunately true that in the struggle to have "my class" come out on top the teacher has almost forgotten the primary purpose of education, to help the learner. In these schools test results have tended to become an end in themselves. When this occurs, "good" test results may be accompanied by poor education.

*Has teacher emphasis been on improving the test results of those just below the median even at the expense of the highest and lowest learners?* This question is closely related to the preceding ones. If a teacher knows that her group and her teaching are likely to be evaluated on the score of the middle learner (median score) of her group, she may gear her teaching to the level of the middle third of the group, feeling that the highest third will get fairly good scores anyhow and that the lowest third or at least the lowest fifth does not have much chance of getting up to the median anyway. Those who are familiar with the percentile system can investigate this possibility by seeing how the 10th, 25th, 75th, and 90th percentiles of the group compare with the norms for each of these levels.

*What is the length of the school term in the system under consideration as compared with the lengths of school terms of the schools on which the norms were based? What would a similar comparison of the annual amount spent per child show?* The time spent with the child and the wherewithal for educating him should be considered in evaluating and interpreting the test results.

*How many years has the test or another form of it been given in the school?* Undue familiarity with a particular test will tend to produce artificially high results on that test. For example, the writer is familiar with a school which twice a year has used some form of the same test for five years, which means that most of the seventh graders in that school have had ten very similar tests and some repeats of the same test. Too great familiarity with a test will indicate the need for using a different test.

*Are test results in certain learnings being achieved at the expense of other equally important learnings?* The basic importance of this point cannot be over-emphasized. The writer has known schools where it appeared that the teaching was so slanted by the standardized tests to be given that many learnings considered at least equally important by outstanding educators were at best neglected. In some schools it even appeared that some desirable attitudes and practices the learners already had were being eliminated to a significant degree. For this reason it is very significant for the interpreter of test results to consider the probable effect of the type of training being used on attitudes toward: acquiring more learning, teachers in general, schools, testing itself, specific subjects, such as art, music, speech. We would also need to consider whether the learner is developing the ability to use the learnings being acquired in practical situations. Is training being given so that there is likely to be much carry-over to out-of-school situations?

*What effect is the method of teaching having on self-adjustment and social adjustment?* This question is closely related to the preceding one, but it is so important that it is given separate consideration. What doth it profit a school to over-emphasize to students information which enables them to pass

an achievement test if thereby the chances of a high degree of self- and social adjustment are decreased? Certainly interpretation of achievement test results should not neglect a consideration of learner adjustment. Personality or adjustment inventories or questionnaires if properly approached by teachers and learners will usually give some clues as to the extent of maladjustment.

*What are the primary motivators of the learners? Are secondary motivators and even undesirable motivators being used to a large degree?* Threats, fear of punishment, marks, extrinsic rewards and the like can produce what seem to be fairly good temporary results but the net result is likely to be undesirable in the long run. To what extent is learning likely to go on without the immediate stimulation of the teacher? Are the learners being pushed too much for their own good? Is the motivation for the learning coming largely from the inside of the learner or from the outside?

*To what extent is it likely that good home teaching and much travel and other educational influences outside of the control of the school have contributed to the high median score on the tests?* Some schools take credit for doing a lot more than they are responsible for, particularly if it is something worthwhile. Schools with learners coming from homes with good "educational opportunities" should give the home credit for many learnings which the pupils demonstrate.

*What is tested by the test? What are the goals of a "good" school? How many goals of a "good" school are tested by the test in question?* These questions are fairly self-explanatory. For example, does the test sample application as well as memorization? Does it contain many items which can be answered by intelligence alone? On what basis has the test been validated? What other tests are available for checking on other important goals of a "good" school?

*Was the test given and scored in a standardized fashion?* It has been found in some schools that "standardized" tests are given in a very unstandardized fashion either through lack of knowledge of good testing procedures or because the person giving or scoring the test has something to gain or to lose by

good or poor performance showing up on the test. If lack of knowledge of good testing procedures is the difficulty, training in this is, of course, indicated. If tests are intentionally given or scored in an unstandardized fashion then the reasons for having the whole testing program needs to be carefully examined.

*What is the potential ability of the learners as compared with those upon whom the tests were standardized?* While this probably cannot be determined exactly, some clues to the answer can be obtained through the use of a good intelligence or mental test. The occupational level of the parents will frequently give a very rough indication of the answer also.

*What are actual living conditions in the community? Is the school teaching the things that are most needed in this community?* These questions are ultimately the most important ones of all that have been raised. No school should be self-satisfied unless satisfactory answers can be given to the following questions suggested in part by Thorndike's *American Cities and States*. What is the infant death-rate? What is the general death-rate? What is the per capita expenditure for recreation? How rare is poverty in the community? What is the excess of physicians, nurses, and teachers over male domestic servants? What is the per capita circulation of the better magazines in the community? What is the average real wage in the community? To use Thorndike's<sup>1</sup> phraseology: What is "the general goodness of life for good people in the community in question, the per capita income of its residents, and their personal qualities of intelligence, morality and care for their families?"

---

<sup>1</sup> Thorndike, Edward L. *American Cities and States: Variation and Correlation in Institutions, Activities, and the Personal Qualities of the Residents*. New York: The New York Academy of Science, 1939.



# STUDIES OF POPULARITY IN COLLEGE: I. CAN POPULARITY OF FRESHMEN BE PREDICTED?

JEAN WAID REILLY AND FRANCIS P. ROBINSON  
Ohio State University

THE typical student entering college wants to have friends, wishes to be popular, and to belong to some well-liked campus group. Students' feverish participation in "rushing" and "ice breaker" activities and the disappointed torpor of the unbid isolated student indicate intense emotional interest. Keen observers have judged that this desire for acceptance and its corollary, the sex drive, are the primary motivating forces of the adolescent age. Counselors, in turn, are interested in fostering students' cordial acceptance of each other not only because school work might otherwise suffer but also because such social participation leads to wholesome personal development.

As the freshman class arrives on the scene, the interested counselor naturally wonders which students will have friends and become leaders and which ones will tend to drift into shy and frustrating isolation. Can the counselor, as he looks over the usual entrance records, make any initial judgments as to the probable future social acceptance of a student? Such predictions would enable the counselor at the time of room assignments, "get-acquainted" parties, and course scheduling to adjust situations in various ways so as possibly to aid the potentially isolated student.

This report presents evidence as to the efficiency of the usual college entrance data in predicting popularity at the middle of the freshman year. To forestall later disappointment it seems best to state here that the predictive efficiency of such data is very low. Many likely items turn out to be

quite valueless and the few predictors found would not at first thought seem to have value. The report does indicate that colleges are failing to obtain important types of entrance data which they should have.

As a first step in this experiment, a measure of social acceptance was obtained for each of a large number of college students. One hundred and sixty-three girls (mostly freshmen) living in a university dormitory in which one of the writers was a graduate counselor, filled out a sociometric test. In this test each girl listed (for each situation) the two or three girls on the campus with whom she would most like "to eat meals," "to study," "to spend my free time," "to attend a football or basketball game," "to room," "to be in a 'bull session,'" "to double date," and "to work on a committee." A popularity score was obtained for each student by tabulating the number of times her name was mentioned by others; a simple weighting scheme was used to give first-choice names more value than those listed second or third. The fact that 100% of the tests were filled out is indicative of the motivation and completeness of the sample. While each girl's score expressed only her popularity in the dormitory group being studied, this particular group seemed representative of the campus population (both sorority and non-sorority), and students usually like most to be accepted in the place in which they have elected to live. (Actual results showed that sorority girls in the dormitory were chosen slightly more frequently than non-sorority girls.)

Popularity scores varied from zero to one hundred and twenty-seven. The distribution of these scores was skewed as these distributions usually are; a few individuals made very high scores and most individuals made below-average scores. Other studies have shown that such scores are highly reliable.

The second step was to obtain from entrance records any data which might seem predictive of later popularity. In all, the following information was collected: the number of children in the family, their chronological age, the number of activities in high school, the number of offices held in high school, the number of scholastic honors in high school, whether the student had decided on a vocational choice, whether her parents were

living or not, the education of the parents, the profession of the father, her religious affiliation, and the size of the home town. Since certain other data are usually obtained during Freshman Week, a study was also made of the relationship of intelligence and the college curriculum selected to later popularity.

The relationship of each of these sets of data to later popularity was analyzed in one of two ways. Where it was possible a comparison was made between the average scores on an entrance variable of the very popular (top third) and of the quite unpopular (lowest third). For those variables not permitting this approach, i.e., parents living or not, the procedure was reversed; thus the average popularity score of students whose parents were living was compared with the average of

TABLE 1  
*The Average Score on Differential Background Factors for  
Each Third in Popularity*

Factor	Unpopular third	Middle third	Popular third	C.R. for greatest difference
1. Intelligence .....	72.7	70.2	69.9	.64
2. Children in family ....	1.50	1.74	1.51	.96
3. Age .....	18.7	18.1	18.00	4.24
4. H.S. Activities .....	6.62	6.57	7.37	1.19
5. Offices in H.S. ....	1.49	1.86	2.02	1.36
6. H.S. honors .....	1.20	1.07	1.17	.41

girls whose parents were not living (one or both). In each case, the statistical significance of the differences found was computed; when the critical ratios exceed 3.00 (standard error of the difference), it can be said with some assurance that that particular trait is predictive of later popularity.

Table 1 summarizes the results for six of the background factors studied. Except for chronological age, none of the factors is predictive of later popularity. It may seem surprising that high-school activities and honors are not predictive but academic popularity either is not related to social popularity or the new opportunities at college allow a restructuring of social relationships. In this highly select group, differences in intelligence seem unimportant; the slight tendency for the more intelligent to be less popular is not statistically significant. A

group entering college is very homogeneous as to age but small differences in age seem to be very important. It is probable, however, that the girls are not reacting to students of nineteen and twenty as "too old"; another explanation seems more probable. Entering freshmen of seventeen or under almost surely came directly from high school to college; their average acceptance score was 47.0. Many entering freshmen of eighteen may have stayed out of school a while or have taken longer to finish, thus losing track of their friends; their average acceptance score was 39.1. Those nineteen and over almost certainly varied from the academic lockstep; these persons not only lost track of their friends but in some instances the factors which made for academic retardation may also have led to retarded social acceptance. The average acceptance score for these older freshmen was 32.8. The normally older sophomores, juniors, and seniors have had an opportunity in college to become acquainted; their average acceptance scores for the different ages showed smaller differences.

The results for the other background factors were as follows:

(7) A girl who has lost one or both parents seems to be at a disadvantage in becoming popular. Those girls with both parents living had an average popularity score of 40.9, those with one or both parents dead scored only 25.9; this difference has a critical ratio of 3.06. An investigation showed as one likely explanation that the eleven girls with a loss of parent had to work for support and did not have as much time for dormitory activities.

(8) Whether the parents had attended college or not did not seem to affect later social acceptance of their daughters. Daughters of fathers attending college scored 39.6, while daughters of non-college fathers scored 40.2. Daughters of mothers attending college scored 42.7, while daughters of non-college mothers scored 38.6. Neither difference is statistically significant, the critical ratio in favor of college-trained mothers being only 1.19.

(9) The profession of the father was not of importance, at least in this instance; this is a heartening manifestation of democracy. The average scores for the various occupational groups were as follows:

Professional and semi-professional . . . .	39.5
Managerial . . . . .	41.1
Sales and clerical . . . . .	39.1
Agriculture, forestry and fishing . . . . .	34.4
Skilled workers . . . . .	37.7

The critical ratio for the greatest of these differences was only 1.03.

(10) The religious affiliation likewise seemed unimportant. The twelve Catholic students had an average score of 41.8, the twenty-seven Jewish students a score of 45.5, and the ninety-eight Protestant students a score of 36.7. Due to an extreme range of popularity scores within each religious group, even the greatest of these differences was not statistically significant, i.e., C.R. = 1.75.

(11) The size of the girl's home town was not significantly related to college popularity. The girls were classified into five groups according to the size of their home towns. The average popularity scores for each of these groups of girls was as follows:

Country and village . . . . .	42.9
Towns under 1,000 . . . . .	37.5
Towns between 1-10,000 . . . . .	38.8
Cities between 10-100,000 . . . . .	41.9
Cities over 100,000 . . . . .	38.7

(12 & 13) The final two items studied were expressions of vocational interest as the girl entered the university, i.e., had she decided on a vocation, and in what college was she enrolled? Neither item was of significance in predicting popularity. The critical ratio in favor of having made a vocational decision was 0.49. The descending sequence of popularity by colleges was Commerce, Arts, Agriculture, and Education, but the critical ratio for the greatest difference was only 1.78.

Since having friends and group status are so important to a student's happiness and development, colleges should give consideration to students' social adjustment. The data in this study show an extreme range among girls in popularity, but the most frequent category is relative social isolation. It is also evident that colleges have little information with which to work as they "arrange the college program to fit the needs of the

entering freshman." In fact, the only two variables usually listed in college entrance data which were related to later popularity (lack of it) were over-age freshman (19 or older) and loss of a parent.

Academic census data needs to be supplemented (and even supplanted if the blank is crowded) by more *vital* statistics from the adolescent world. Thus the writers hazard the guess that similar informal sociometric results for the high-school senior class, simple counts of who is chosen for committees or for partners in games, and whose first name or nickname is known by the most students might be predictive of later popularity in college or in adult life.

## MEASUREMENT ABSTRACTS<sup>1</sup>

Anderson, John E. "Parents' Attitudes on Child Behavior: A Report of Three Studies" *Child Development*, XVII (1946), 91-97.

A study by Marion J. Radke on the relation of parental authority to children's behavior and attitudes, and one by Katherine A. Miles on the relationship between home background factors and qualities of leadership in children, have furnished techniques for a third study by Frank Hansen, now in process of completion. Generalizations tentatively made on the basis of the studies are: (1) The child forms an image of his parents which tends to emphasize disciplinary and management controls. (2) The parents' attitudes with respect to child management are related to the behavior of the child in his own social group. (3) The parents' concept of the ideal child is related to the behavior of the child in his own social group. (4) There is a relation between children's behavior in social groups and scores on personality measures, and between parents' attitudes toward children and scores on personality measures. *Frances Smith.*

---

Brogden, Hubert E. "An Approach to the Problem of Differential Prediction." *Psychometrika*, XI (1946), 139-154.

A procedure for maximizing selective efficiency is developed for application to situations in which it is desired to select from a single group of applicants for several possible assignments. The problem of comparable units for the several criteria whose values must be compared to each other for differential assignment purposes is discussed. It is demonstrated that, assuming linear regressions, maximal selection is obtained if individuals in any given assignment are differentiated from those rejected according to critical rejection scores on the multiple-weighted sum of the predictors and from another possible assignment by critical difference scores which are merely the differences between the two critical rejection scores. Since the relationships just indicated give no way of determining the magnitude of the critical scores required to select the required number of persons for each assignment, a successive approximation procedure for accomplishing this purpose has been devised and a computational example is worked out. (Courtesy *Psychometrika*.)

---

<sup>1</sup> Edited by Forrest A. Kingsbury.

Christensen, Thomas E. "Some Observations with Respect to the Kuder Preference Record." *Journal of Educational Research*, XL (1946), 96-107.

The author administered the *Kuder Preference Record* to twenty-seven ninth-grade pupils to determine (1) how the meanings of items on the Kuder are interpreted, (2) whether systematic class instruction on the meaning of items affected the reliability of the test, and (3) the extent to which chance influences scores. It was discovered that because of a lack of occupational information, the children did not comprehend the meaning of many of the items, invalidating many of the scores. Systematic class instruction did change scores, and when preference was determined entirely by chance, high scores could be obtained and interpreted as significant when they were not actually so. *Harold Mosak.*

---

Despert, J. Louise and Pierce, Helen Oexle. "The Relation of Emotional Adjustment to Intellectual Function." *Genetic Psychology Monographs*, XXXIV (1946), 3-56.

Thirty-nine normal children were subjected to intensive personality studies during their two or three years' attendance at the Payne Whitney Nursery School of New York City during the period 1937-1942, the purpose being to discover any relationship between intellectual function and emotional adjustment. The investigation included anamnestic data, physical examinations, psychometric test data, and individual records of daily behavior and play sessions. Significant changes in IQ for twenty-two of the children (twelve with increases and ten with decreases) indicated a close parallel between psychometric findings and emotional adjustment as judged from all the data. While advancing no hypothesis for this influencing of intellectual function by emotional causes, the authors believe the study shows a need for supplementing psychometric testing with a projective technique for detecting these causes. *Vernon S. Tracht.*

---

Eysenck, H. J. and Himmerweit, H. T. "An Experimental Study of the Reactions of Neurotics to Experiences of Success and Failure." *Journal of General Psychology*, XXXV (1946), 59-75.

Two groups of 100 male neurotic army patients, equated for age and intelligence, were given a modified form of the *Pursuit Rotor* to obtain their performance scores, aspiration scores, and judgments of past performances. The results showed no significant differences for performance or improvement on the test between the fifty with predominantly hysterical symptoms and the other fifty with largely affective symptoms. However, the latter group were significantly different from the former in having greater deviations from their actual scores with regard to the level of aspiration and judgment of past performance. In general, the reactions to success and failure of the affective group were more subjective as compared with the greater objectivity of the hysterical group, thus tending to confirm



Jung's analysis of extraverted (hysterical) and introverted (affective) personality traits. *Vernon S. Tracht.*

---

Gross, Sister M Mynette. "The Effect of Certain Types of Motivation on the 'Honesty' of Children" *Journal of Educational Research*, XXX (1946), 133-140.

To determine the effect of self-competition and group-competition on "honesty," 229 seventh-grade children were divided into a control group and two experimental groups. Honesty was measured by the self-scoring of the *Clapp-Young Arithmetic Test* given each day for a week. One experimental group was encouraged in group-competition, and the other in self-competition. Results indicate a wide variation in the number of children who changed their answers in all three groups. The type of motivation did not increase "dishonesty." Individual consistency from day to day in the number of answers changed was low. The mean IQ of the "honest" children was higher than for the "dishonest" children. *Betty Steele*

---

Harris, Adeline and Watson, Goodwin. "Are Jewish or Gentile Children More Clannish?" *Journal of Social Psychology*, XXIV (1946), 71-76.

Eighty-two children in Grades IV-VI of a private school were asked to list the names of their best friends among children in their own class, among those in the whole school, and among those seen only outside school. The study shows that Gentile children made a larger percentage of in-group choices than did the Jewish in all three categories, in-group choices being made most frequently by both groups outside of school. Gentile girls made more in-group choices than did Gentile boys, while among the Jewish children the boys made the larger number of in-group choices. There was no evidence of an increase or decrease in exclusiveness within the limits of the grades studied. One-fourth of the Gentile children had no Jewish friends, but all of the Jewish children had some Gentile friends. *Frances Smith.*

---

Hayes, Samuel P., Jr. "Diagrams for Computing Tetrachoric Correlation Coefficients from Percentage Differences." *Psychometrika*, XI (1946), 163-172.

A description is given of diagrams (available separately) for computing tetrachoric correlation coefficients. The diagrams are entered with "per cent of combined groups above dividing point" and the difference between groups in their per cents above the dividing point. (Courtesy *Psychometrika*.)

---

Jones, Ronald Devall. "The Prediction of Teaching Efficiency from Objective Measures." *Journal of Experimental Education*, XV (1946), 85-99.

This article presents the results of a study of sixty-five University of Wisconsin trained teachers to determine how well objective data

can predict teaching efficiency as measured by the criteria of supervisory ratings and pupil-gain scores. The objective data employed included rank in high-school class, university grade averages, grades in three education courses, the *Henmon-Nelson Test of Mental Ability*, the *Bell Adjustment Inventory*, and the *Link Interest Inventory*. The results demonstrate that the criteria are not related to each other to a degree greater than can be attributed to chance. High-school rank is apparently the best predictive measure of residual pupil gain. *Harold Mosak.*

---

Lawshe, C. H., Semanek, Irene A., and Tiffin, Joseph. "The Purdue Mechanical Adaptability Test." *Journal of Applied Psychology*, XXX (1946), 442-453.

Two forms of the *Purdue Mechanical Adaptability Test* were administered to industrial applicants and employees. After eliminating those items associated with mental ability, sixty items were retained in a revision, Form A. Split-half reliability of the revised test was  $.84 \pm .01$ . Validation studies were made employing job success as estimated by supervisor's ratings as the criterion. Form A correlates  $.71 \pm .09$  with the *Bennett Test of Mechanical Comprehension*, but it is apparently little related with intelligence since its correlation with the *Otis Self-Administering Higher Examination* is only  $.08 \pm .20$ . These results lead the authors to the conclusion that the test "is useful in identifying men or boys who are mechanically inclined and are likely to succeed in jobs of a mechanical nature. . . ." *Harold Mosak.*

---

Lins, Leo J. "The Prediction of Teaching Efficiency." *Journal of Experimental Education*, XV (1946), 2-60.

The purpose of this study was to determine whether any material in the pre-service records and personal make-up of applicants could be employed in predicting (1) whether the person would be admitted to the School of Education at the end of the junior year, and (2) subsequent teaching efficiency. The data studied included replies to the University of Wisconsin application blank, intelligence test scores, autobiographical material, and academic achievement. The author finds that the best criterion for predicting admission to the School of Education is a composite of freshman-sophomore grade-point averages. High-school rank and grades in practice teaching were among those items with a predictive value for teaching efficiency. *Harold Mosak.*

---

McMillan, Robert T. "School Acceleration and Retardation Among Children in Southern Oklahoma." *Journal of Educational Research*, XL (1946), 126-132.

One group of 172 children from grades I through XII were selected from a village of 700 population in southwestern Oklahoma, and a second group of 150 children also from grades I through XII were selected from a near-random sampling in seventeen villages in southeastern Oklahoma. The groups were compared in terms of school

acceleration and retardation based on the expectation of completion of one grade per year of attendance. Results indicated that acceleration and retardation is influenced by the family socio-economic status, the schooling of parents, the amount of moving done by parents, the number of children in the family, and then sex and age. *Betty Steele.*

---

Mandell, Milton M. "The Group Oral Performance Test." *Public Personnel Review*, VII (1946), 209-212.

The *Group Oral Performance Test* now in use by the British Civil Service Commission for the selection of administrative and foreign service personnel is superior to the oral interview method commonly used by civil service commissions in that it provides evidence on the individual's behavior in a group situation. It also depends less on the individual rater's ability to "draw out" the candidate; it may be used to reproduce the work situation; and it is more economical in time expenditure. Testing techniques developed in the trial of this method by the United States Civil Service Commission are described, and a sample group task for candidates is presented. *Frances Smith.*

---

Manson, Morse P. and Grayson, Harry M. "Keysort Method of Scoring the Minnesota Multiphasic Personality Inventory." *Journal of Applied Psychology*, XXX (1946), 509-516.

This gives a detailed description of a method of surmounting a practical difficulty in the use of the MMPI, i.e., the lengthy scoring technique. As developed and used in the Personnel Evaluation Department of the Army's Disciplinary Training Center in the Mediterranean Theater of Operations, such an improved scoring method minimizes clerical error, eliminates the recording of each separate item, and shortens the scoring time by more than 15 minutes. Adapted from the McBee Keysort System, the 550 MMPI cards (containing various statements about attitudes, interests, and general behavior of the testee) are notched in a like manner to facilitate the separation of items into the three principal categories (deviate, *no* deviate, and *no* non-deviate), by inserting a rod or needle through the proper holes in the stack of cards. Four needle movements give the sorting of the cards necessary for scoring by simple count each of the twelve subtests comprising the total inventory. *Vernon S. Tracht.*

---

Meehl, Paul E. "An Investigation of a General Normality or Control Factor in Personality Testing." *Psychology Monographs*, L (1946), 1-62.

A scale of seventy-eight items was empirically derived from the total item pool of the *Minnesota Multiphasic Personality Inventory* in an attempt to quantify Rosanoff's hypothetical factor of general "normality" or "control." Based on its ability to differentiate between a criterion group of apparently normal persons with marked deviations on the "neurotic triad" of the MMPI and a matched group of clinically abnormal persons with similar scores, this so-called *N*

scale had fairly good reliability and proved valid when used on a new test group of normals. Suggested interpretations as to the psychological nature of  $N$  are briefly discussed, the one deemed most feasible being that it probably measures a "plus-getting" tendency inherent in personality tests generally as roughly related to the tendency of certain persons to obtain spurious psychasthenic scores on the MMPI; and recommendations for further research are advanced. *Vernon S. Tracht.*

---

Meehl, Paul E. "Profile Analysis of the Minnesota Multiphasic Personality Inventory in Differential Diagnosis." *Journal of Applied Psychology*, XXX (1946), 517-523.

To evaluate the MMPI as a differential diagnostic instrument employing a rapid, inspectional technique of pattern analysis of profiles, the records of 147 hospitalized psychiatric patients were subjected to a procedure described as "blind" diagnosis. Clinical appraisal by the psychiatric staff, made before the present MMPI scales (with one exception) were developed, constituted the criterion. Of the approximately  $2/3$  actual abnormals correctly identified, about  $2/3$  were placed in the appropriate three major categories—psychosis, psychoneurosis, and conduct disorder—the contingency coefficient for agreement between "blind" and actual diagnostic groupings being .55. Although such discriminations, among the study's other findings, are statistically much better than chance, the author admits that the proportion of false classifications is considerable but expects the margin of error to be reduced by more mathematically precise use of the suppressor  $K$  and by greater formalization of pattern analysis. *Vernon S. Tracht.*

---

Meehl, Paul E. and Hathaway, Starke R. "The  $K$  Factor as a Suppressor Variable in the Minnesota Multiphasic Personality Inventory." *Journal of Applied Psychology*, XXX (1946), 525-564.

This article discusses the general problem of the effect of test-taking attitudes on scores derived from structured personality inventories. After surveying briefly the literature on the subject and the various approaches previously made toward solving this problem, it tells of attempts to construct special scales measuring these attitudes as applied specifically to subjects taking the MMPI. The use of one such scale, embodying the so-called  $K$  variable, is shown to be relatively successful in detecting the presence of disturbing test-taking attitudes, thereby improving the means of differentiating normals from abnormals. *Vernon S. Tracht.*

---

Mitchell, Claude. "How Valid Are Vocational Analysis Blanks." *Journal of Educational Research*, XL (1946), 57-62.

In 1927 a senior high-school class was given the *Vocational Guidance Questionnaire* and the *Choices of Working Conditions Blank*. In 1945 the questionnaire information was checked against the actual vocational status of fifty-three subjects. Results showed that 49 per

cent followed vocations selected eighteen years previously on the questionnaire. The results of the *Choices of Working Conditions Blank* suggest that failure to follow occupational preferences is caused by the failure to secure the preferred work and not by a change of interest. These results indicate that self-analysis blanks have some validity. *Betty Steele.*

---

Patterson, Cecil H., Jr. "The Relationship of Bernreuter Scores to Parent Behavior, Child Behavior, Urban-Rural Residence, and Other Background Factors in 100 Normal Adult Parents." *Journal of Social Psychology*, XXIV (1946), 3-49.

The author attempts in this study to determine whether there exists any justification for utilizing the *Bernreuter Personality Inventory*. He employed this and other measures with one hundred married couples to determine whether any significant relationships existed between the scores obtained on the Bernreuter and such factors as urban-rural residence, parent and child behavior, and age at, and duration of, marriage. He concludes that although the Bernreuter is reliable, it possesses little validity since other studies find a significant relationship between the personalities of parents and children and the Bernreuter fails to do so. He maintains further that the use of all six scorings on the Bernreuter is wasteful since they are all highly intercorrelated. *Harold Mosak.*

---

Rabin, A. I., Davis, J. C., and Sanderson, M. H. "Item Difficulty of Some Wechsler-Bellevue Subtests." *Journal of Applied Psychology*, XXX (1946), 493-500.

On the basis of the individual Wechsler-Bellevue records of 300 normal individuals, an analysis was made of item difficulty in six subtests of the scales: Information, Comprehension, Similarities, Picture Completion, Picture Arrangement, and Block Designs. Items within each test were ranked according to the number of subjects passing each one. The results indicate that revisions in order of presentation are desirable in all of the subtests but the Picture Arrangement, tests of Information and Comprehension in particular showing variation from the order suggested in the third edition of Wechsler's *Manual*. Suggested changes in the original order of item presentation are summarized in table form. Findings on 1000 psychiatric patients substantiate the study. *Frances Smith.*

---

Schlesser, George E. "Development of Special Abilities at the Junior High-School Age." *Journal of Educational Research*, XL (1946), 39-51.

The present investigation of special abilities included the problems of their permanence, their related factors, the size of their significant deviations, and the relative contributions to achievement of special and general abilities. The *Unit Scales of Attainment* battery was given to 104 sixth-grade and junior high-school students who previously had been given the *Standard Achievement Test* in the third

grade. The results found were: (1) A special ability is one which is at least 1.8 school years advanced beyond the general level of ability. (2) Special abilities tend to change rather than to become stabilized, depending on test reliability. (3) The pupil's estimate of his profile is poor compared to the actual profile. (4) The profile is not made more even or uneven during junior high school. (5) Special abilities account for 45 per cent of the total variance in achievement, and general abilities for 55 per cent. *Betty Steele.*

---

Springer, N. Norton. "The Validity of the Multiple Choice Group Rorschach Test in the Screening of Naval Personnel." *Journal of General Psychology*, XXXV (1946), 27-32.

To determine its value as a psychiatric screening device in military selection, the author gave the *Harrower-Erickson Multiple Choice Group Rorschach* to 471 naval personnel in service sixteen months or more. Of these, 217 had good duty status and were in this country awaiting further assignment, while the remaining 254 had bad conduct records and were either serving court martial sentences or had been discharged after the completion of sentences. He found no reliable difference between those making a good adjustment and the chronic offenders to naval discipline, thereby concluding that this type of Rorschach in its present form fails to differentiate normals from those classified as having various behavior disorders, and that it is therefore unsuitable for such purposes. *Vernon S. Tracht.*

---

Symonds, Percival M. "Evaluation of Teacher Personality." *Teachers College Record*, XLVIII (1946), 21-34.

This article explains how the newer psychological techniques of personality appraisal can be applied advantageously to the selection of teachers. Among others, the important features of ratings, interviews, stress situations, psychodrama and projective test methods are briefly outlined in the light of four proposed personality factors deemed essential to the successful teacher. Because of the proven inadequacy of older ways based on casual interviews with candidates and superficial observation of their educational records by superintendents, the author proposes personality assessment by a central agency whose findings would be available to all school officials, with the consent of the prospective teachers. Due allowance would be made for those aspects of personality which are changeable, such as interests and motivation. *Vernon S. Tracht.*

---

Thurstone, L. L. "Note on a Reanalysis of Davis' Reading Tests." *Psychometrika*, XI (1946), 185-188.

Correlation data on nine reading tests originally analyzed by Frederick B. Davis by the principal axes method are reanalyzed by Spearman's uni-dimensional method. It is concluded that a single common factor (reading ability) accounts for the correlations among the tests with residuals remarkably small in view of the fact that the

tests were designed to test nine supposedly different skills. Three of the tests showed additional specific variance not attributable to the common factor. (Courtesy *Psychometrika*.)

---

Tuckman, Jacob. "The Relationship Between Subjective Estimates of Personal Adjustment and Ratings on the Bell Adjustment Inventory." *Journal of Applied Psychology*, XXX (1946), 488-492.

The *Bell Adjustment Inventory* and the *Adjustment Questionnaire*, the latter consisting of five self-rating statements based on the Inventory, were given to 191 boys and 200 girls in grades IX-XII, and to fifty-one men and forty-five women referred for vocational guidance. The corrected contingency coefficients, except for the emotional adjustment for boys, were high enough to warrant the substitution of the *Adjustment Questionnaire* for the *Adjustment Inventory*. From 22 per cent (health adjustment for boys) to 44.4 per cent (home adjustment for women) have identical ratings on both questionnaires. From 37.2 per cent to 64.4 per cent fall within the same group of above average, average, and below average in adjustment in both questionnaires. *Betty Steele*.

---

Von Haden, Herbert I. "An Evaluation of Certain Types of Personal Data Employed in the Prediction of Teaching Efficiency." *Journal of Experimental Education*, XV (1946), 61-84.

Because of the current teacher shortage and the consequent exercise of less discrimination in the selection of teaching personnel, the prediction of teaching effectiveness has assumed additional importance. The value of subjective data only, such as instructors' comments, interviews, and autobiographies, has been studied. The author concludes that estimates from these interviews, autobiographies, and instructors' comments are not closely related to efficiency. Of the personal qualities whose influence upon future success in teaching were studied, the most readily defined and evaluated were work habits, initiative, and professional judgment. *Harold Mosak*.

---

Wherry, Robert J. and Gaylord, Richard H. "Test Selection with Integral Gross Score Weights." *Psychometrika*, XI (1946), 173-183.

A new method of test selection, which attempts to combine the merits of the Toops *L*-Method with those of the Wherry-Doolittle Method, is presented. It results in integral (unit if desired) positive and/or negative (optional) weights. This flexibility makes the method applicable to all kinds of material and for selecting items for tests and tests for batteries. An explicit solution of one test construction problem is presented. Necessary changes in method for the solution of five other types of test construction problem are presented. A few cautions are provided for potential users. (Courtesy *Psychometrika*.)

Wherry, Robert J. and Taylor, Erwin K. "The Relation of Multiserial Eta to Other Measures of Correlation." *Psychometrika*, XI (1946), 155-161.

Ordinary product-moment correlation and regression methods are frequently not immediately applicable to qualitative data, whereas multiserial  $r$ , point-multiserial  $r$ , and multiserial eta can be easily applied. The multiserial  $r$  is rejected for prediction since it tells us only what the correlation *might be* if certain assumptions *were* true and if we *could* measure what is *not now* measured. The point-multiserial  $r$  and multiserial eta are identical when the number of categories is two but differ when it is three or greater. The multiserial eta is identical with the product-moment  $r$  when categories are assigned scale values equal to their means on the continuous variable. With three or more categories, the point-multiserial  $r$ , which *assumes* linearity with *equal* step intervals, is always lower than the multiserial eta, which *forces* linearity by adoption of *unequal* step intervals based upon difference in criterion attainment. While the multiserial eta expends one degree of freedom with the weighting of each category, this is known and correctable, whereas the vague partial loss of degrees of freedom due to the ordering of categories in the point-multiserial is not correctable. (Courtesy *Psychometrika*.)

Winfield, Marjorie Case. "The Use of the Harrower-Erickson Multiple Choice Rorschach with a Selected Group of Women in Military Service." *Journal of Applied Psychology*, XXX (1946), 481-487.

As a means of discovering possible misfits among the Marine Corps' enlisted women personnel about to be sent overseas on a voluntary basis, the *Harrower-Erickson Multiple Choice Rorschach* and the *Minnesota Multiphasic Personality Inventory* were given to the first group of 181 to reach the staging area. Neither test's scores, whether used alone or in combination with the other, was considered enough for differential diagnosis or as sufficient grounds for excluding persons from going abroad without verification through systematic observation by staff officers. Results indicated no correspondence between MCR and MMPI scores nor between the MCR and observed behavior; hence the author concludes that the MCR measures something other than it purports to do, and is therefore unsuitable for screening purposes until further research and standardization is done. *Vernon S. Tracht*.

Zangwill, O. L. "Some Qualitative Observations on Verbal Memory in Cases of Cerebral Lesion." *British Journal of Psychology, General Section*, XXXVII (1946), 8-19.

Records of 200 neuropsychiatric cases examined at the Brain Injury Centre, Edinburgh, were analyzed with special reference to disorders of language and of memory as related to three types of verbal test. The test results studied were those for auditory-verbal memory span, rote learning, and substance recall. Impairment of



memory span for verbal material and severe difficulties in rote learning characterized the dysphasic cases, while cases of dysmnnesia were found to have intact verbal memory span but grossly impaired verbal learning. Substance recall, while seriously impaired in the amnesic syndrome, is often well preserved in dysphasic conditions. Implications of the findings for a theory of the relationship of memory to habit are discussed, and their bearing on practical problems of diagnosis and treatment is indicated. *Francis Smith.*

#### ADDITIONAL ARTICLES NOT ABSTRACTED

- Anderson, William Ewart. "An Attempt Through the Use of Experimental Techniques to Determine the Effect of Home Assignments Upon Scholastic Success." *Journal of Educational Research*, XL (1946), 141-143.
- Feder, D. D. "An Approach to Test Analysis for Public Personnel Agencies." *Public Personnel Review*, VII (1946), 126-131.
- Kellogg, W. N. "The Learning Curve for Flying an Airplane." *Journal of Applied Psychology*, XXX (1946), 435-441.
-



## THE CONTRIBUTORS

Steuart Henderson Britt—Ph.D., Yale University, 1935. Fellow and Assistant in Psychology, Institute of Human Relations, Yale University, 1932–1935. Research Assistant, Institute of Educational Research, Columbia University, 1935–1936. Assistant Professor, George Washington University, 1936–1942. Visiting Lecturer in Psychology, University of Virginia, summer 1941. Chairman, Subcommittee on Listing of Personnel in Psychology, 1941–1942; Executive Director of Office of Psychological Personnel, National Research Council, 1942–1943. Secretary of N.R.C. Emergency Committee in Psychology, 1942–1945. Consultant, National Resources Planning Board, 1941–1942. Chief, Military Advisory Section of the National Roster of Scientific and Specialized Personnel, War Manpower Commission, 1942–1943. Executive Secretary, National Resources Planning Board's Committee on Wartime Requirements for Specialized Personnel, 1943. Reserve officer, U.S.N.R., 1943–1946. Consultant to Westinghouse Electric Corporation and Science Service on Annual Science Talent Search, 1942–. Associate Director of Research, McCann-Erickson, Inc., New York, 1946–. Author of *Social Psychology of Modern Life*; co-author of *Jews in a Gentile World*, *The Problem of Anti-Semitism*, and author of articles in learned journals on problems of social psychology, public opinion, legal psychology, and personnel work. Fellow, American Association for the Advancement of Science, American Psychological Association. Member, American Sociological Society, American Bar Association, Society for Psychological Study of Social Issues, National Education Association, American Educational Research Association, American Marketing Association, New York Academy of Science.

Walter N. Durost—Ph.D., Teachers College, Columbia University, 1932. Graduate Assistant in Educational Measurement, 1931–1932; Research Assistant, 1933–1934, Assistant in Educational Research, 1934–1935, Teachers College, Columbia University. Test Editor and Director, Division of Tests and Research, World Book Company, Yonkers-on-the-Hudson, 1935–1942. Personnel Research Technician, U. S. Army Air Forces, A.S.C., 1942–1944. Field Representative on Testing, U. S. Army Service Forces, 1944–1945. Director, Division of Tests and Research, World Book Company, 1945–. Author of articles on tests and measurement. Fellow, American Psychological Association. Member, American Educational Research Association.

Harold A. Edgerton—Ph.D., Ohio State University, 1928. Research Assistant, Mechanical Abilities Investigation, University of Minnesota, 1924–1925. Research Assistant, Department of Psychology, 1928–1930; Assistant Professor of Psychology, 1930–1941; General Supervisor Statistical Unit, Occupational Research Program, U.S.E.S., 1935–1936; Counselor, College of Arts and Sciences, 1937–1941; Director, Occupational Opportunities Service, 1941–1946; Associate Professor of Psychology, 1941–1945, Ohio State University. Director, N.R.C. Research on Development of Criteria for Use in the Study of Selection and Training of Pilots, 1939–1942. Civilian Education Supervisor, Star Unit (1547 S.U.), 1943–1944. Expert Consultant, Personnel Research Section, T.A.G.O., 1944–. Consultant, Science Service, Inc., Washington, D. C., on Westinghouse Annual Science Talent Search. Author of *Academic Progress: A Follow-Up Study of the Freshmen Entering the University in 1923*, *Academic Prognosis*; co-author of *Statistical Dictionary*; and author of articles in psychological and educational journals. Member, Sigma Xi, American Psychological Association, American Association for the Advancement of Science. President, Psychometric Society.

John E. Horrocks—Ph.D., Syracuse University, 1945. Industrial and Y.M.C.A. work, 1930–1935. Teacher and Director of Guidance, New York State Public Schools, 1936–1942. Administrative Assistant, 1942–1944; Instructor in Educational Psychology, 1944–1945, Syracuse University. Assistant Professor of Psychology, Ohio State University, 1945–. Author of standardized industrial and psychological tests, and of articles in various educational and psychological journals. Member, American Psychological Association, American Association for the Advancement of Science, Sigma Xi.

(Mrs.) Jean Waid Reilly—M.A., Ohio State University, 1942. Assistant Personnel Adviser to Women, 1942–1943; Personnel Adviser to Women, 1943–1944, University of North Carolina. Personnel Counselor, Rehabilitation Program, U. S. Naval Hospital, Pensacola, Florida, 1944–1945. Assistant, Student Psychological Consultation Service, Ohio State University, spring 1946. Teaching Assistant, Department of Psychology, Ohio State University, 1946–.

Francis P. Robinson—Ph.D., University of Iowa, 1932. Instructor, University of Iowa, 1932–1933. Chairman, Department of Psychology, Stout Institute, 1933–1937. Assistant Professor of Psychology, University of Oregon, 1937. Assistant Professor of Psychology, 1937–1941; Associate Professor of Psychology, 1941–1944; Professor of Psychology, 1944–, Ohio State University. Author of *Diagnostic and Remedial Techniques for Effective Study*, *Psychology and the New Education*; co-author of *Effective Study*, and author of articles on educational psychology and student personnel. Fellow, American Psychological Association.

Ray H. Simpson—Ph.D., Columbia University, 1938. Lecturer, 1934–1936; Instructor, 1936–1939, Barnard College, Columbia University. Assistant Professor of Psychology, 1939–1942, University of Alabama. Personnel Consultant, U. S. Armed Forces, 1942–1946. Professor of Psychology, 1946–, University of Alabama. Author of *Those Who Influence and Those Who Are Influenced in Discussion*, and of articles in educational and psychological journals.

Harry C. Steinmetz—Cand. for Ph.D., Purdue University, 1947. Associate Professor of Psychology, San Diego State College, 1930–. President, Central Labor Council, A.F.L., San Diego, 1935–1936. Private Clinical Consultant, 1942–. Author of *Industrial Efficiency Rating*; co-author of *Manual of Aircraft Layout* and author of articles in psychological journals. Member of various psychological associations. Vice-President, American Federation of Teachers, 1936–1938. President, San Diego Psychological Society, 1946.

Maurice E. Troyer—Ph.D., Ohio State University, 1935. Superintendent, Bureau of Township Schools, Princeton, Illinois, 1925–1929. Assistant Professor of Psychology, Bluffton College, 1930–1932. Instructor in charge of Remedial Program, Ohio State University, 1933–1936. Assistant Professor of Education, 1936–1939; Associate Professor, 1939, Syracuse University. Associate in Evaluation, Commission on Teacher Education, American Council on Education, 1940–1943. Director, Bureau of School Services; Professor of Education, 1943; Director, Evaluation Service Center, 1945, Syracuse University. Member, American Psychological Association, American Association of Applied Psychology, American Educational Research Association, American Association for the Advancement of Science.



# The American College Personnel Association Supplement

Edited by  
FRED McKINNEY  
University of Missouri

1947





# THE AMERICAN COLLEGE PERSONNEL ASSOCIATION SUPPLEMENT

## CONTENTS

<i>Editor's Note</i> .....	92
<i>The GED Tests as Predictors of Scholastic Success.</i> ROBERT CALLIS and C. GILBERT WRENN .....	93
<i>An Analysis of Four Orally Administered Measures of Adjust- ment.</i> WILLIAM D. ALTUS and HUGH M. BELL .....	101
<i>The Role of the Psychologist in the Counseling Program of the Veterans Administration.</i> GWENDOLEN SCHNEIDLER .....	117
<i>Community Relationships in Out-of-School Counseling.</i> WAR- REN K. LAYTON .....	127
<i>Estimation of Test Item Difficulty by Averaging Highest and Lowest Quarter Performance Compared with Total Population Count.</i> DANIEL D. FEDER ..	133
<i>Application of Informal Projective Methods in the Counseling Interview.</i> HAROLD B. PEPINSKY .....	135
<i>Counseling and the Minnesota Point of View.</i> E. G. WILLIAM- SON .....	141
<i>Social Adjustment and Interest Scores of Introverts and Extro- verts.</i> M. CATHARINE EVANS .....	157
<i>Members of the American College Personnel Association</i> .....	169
<i>The Contributors</i> ....	185

## EDITOR'S NOTE

It has been the tradition of the American College Personnel Association to publish a report of its annual meeting which contained the discussion and research of its members on personnel problems. The national emergency limited meetings and publications, but college personnel workers have been actively attacking theoretical and practical problems related to their activities. This section is a report of some of these research activities. Certain of these papers were presented at recent meetings, others represent the current research and thinking of ACPA members. This collection of writings is wide in scope as are the interests, training and background of college personnel officers. The contributions herein cover differing techniques, GED tests, orally administered personality tests, interest inventories and projective devices; they include discussion of personnel programs in perspective and in practice for unselected students, veterans and former students; one paper deals with test item difficulty.

Any collection of publications in this field reminds us how great is the need for research. No member of the college roster is busier than one who deals with the total needs of the individual student and yet no academician has greater need or desire to empirically check his activities by accumulated records and statistical treatment thereof. May the future provide materials for many sections such as this.

Fred McKinney

# THE GED TESTS AS PREDICTORS OF SCHOLASTIC SUCCESS

ROBERT CALLIS AND C GILBERT WRENN

University of Minnesota

THE prediction of scholastic success has commanded the attention of psychologists and educators for several years, with much research being done on this subject. More recently this question of scholastic success has been applied to veterans. Certain speculations have been made concerning the motivation and maturation peculiar to veterans and the relation of these to scholastic success. Now that veterans are beginning to return to the campuses, it is desirable to be able to predict the scholastic success of these veterans in order that the colleges and universities may be able to counsel with them and to help them efficiently plan their college training. To date, the primary use that has been made of the General Educational Development Tests is in determining advanced standing for veterans; however, it is conceivable that these tests are efficient predictors of scholastic success. The purpose of this study was to determine the extent to which these tests would predict the scholastic success of veterans. Specifically the problem may be stated—*To what extent will the GED Tests predict the honor-point ratio of veterans who are first-term freshmen in the General College of the University of Minnesota?* This problem can best be understood in the light of previous studies of the prediction of scholastic success.

## *Prediction of Scholastic Success*

Segel (6) summarized the research to 1934 on this problem, and Eurich and Cain (5, 846-853) in 1941 again summarized the research. High-school rank was found to be the best single predictor of scholastic success in college, with intelligence test

scores a close second, but the predictive power of these measures varied widely from population to population. The median correlation coefficients of these various studies between college marks and high-school rank or intelligence test scores falls between 0.50 and 0.55. These findings may not necessarily hold true for veterans, since veterans may differ from regular college students in matters of motivation, maturation, and experience. No studies are available concerning the scholastic success of veterans as such; however, there have been a few studies made with the GED Tests on veterans and non-veterans.

### *Previous Findings Concerning the GED Tests*

The study of Yale students by Crawford and Burnham (2) was designed to determine the value of the GED Tests scores in predicting the scholastic success of freshmen. A representative sampling of the entire class, veterans and non-veterans, was used in this study with scores of the *College Entrance Examination Board Tests* used to select the sample.<sup>1</sup> Total standard scores on the GED Tests were found to correlate with first-term freshmen's marks with a coefficient of 0.56 as compared to a coefficient of 0.53 between CEEB total scores and first-term freshmen's marks. This coefficient (0.56) was higher than that for any of the GED Tests used singly. Table 1 shows these relationships.

The study of the GED Tests scores of 114 Harvard students

TABLE 1

*The Relationship of GED Tests Scores to Average First Term Marks for Yale Freshmen, N=135*

	Standard scores		
	M	S.D.	r
GED Total .....	270.4	20.9	0.56
GED I .....	62.9	6.2	0.51
GED II .....	69.5	7.7	0.50
GED III .....	73.3	5.5	0.36
GED IV .....	64.7	6.1	0.41

<sup>1</sup> Other contributions to the present limited knowledge of these tests were provided in this study. The intercorrelations between the various tests were determined and are reported as varying between 0.47 and 0.63 ( $N=135$ ). The total standard scores of the GED Tests correlated well ( $r=0.72$ ) with CEEB Tests total scores.

by Dyer (3) indicated that the tests had small value in determining the advanced standing of students but that they could be useful for admission purposes. For the sample the total standard scores on the GED Tests correlated with college rank (based on marks) as follows:

Non-scientific group .....	0.41
Scientific group .....	0.52
Total group .....	0.46

The GED Test II (Social Studies) predicted scholastic success for the scientific group better than it did for the non-scientific group.

Bradley's (1) study of GED Test II (Social Studies) scores of 100 MacMurray students indicated that these scores correlated with marks in social science courses with a coefficient of 0.66.<sup>2</sup>

#### *Definition of the Sample*

The study reported here was based upon veteran students in the General College of the University of Minnesota. Veterans who wished to apply for advanced standing had taken the GED Tests but it was necessary to limit the sample by applying other criteria to the selection. The total criteria for the group selected were:

1. Was a veteran of World War II;
2. Had entered the General College in Fall Term, 1945, or Winter, 1946, and was in attendance during the Winter Term, 1946;
3. Had no previous formal college training;
4. Had taken the GED Tests and had applied for advanced standing on the basis of these tests;
5. Had completed one term in General College so that marks were available for this term's work.

One hundred and forty-three veterans met all of the criteria but No. 4 and when this was applied the group was reduced to 60. This provided the basic group for the study, although

---

<sup>2</sup> Other data presented in this study indicated that there was no significant relationship between the scores on this test and the number of credits earned in the social science field.

the incompleteness of certain items of information used reduced the number to below 60 for some phases of the study. A population of 60 is small but adequate if it is representative in significant characteristics of the total veteran population. A test of scholastic aptitude (*Ohio State University Psychological Examination*, form 21) was used to determine the degree of representativeness of the sample.<sup>3</sup> It was found that there was no significant difference between the sample and the group of 143 veterans, the critical ratio being 0.59. A further test of representativeness was made by comparing the sample with the total General College population,<sup>4</sup> the critical ratio between these two groups being 2.42 with the veterans group having the higher average score. This difference is significant at the 2% confidence level. On the basis of data presented here, it can be reasonably concluded that the veterans group used is representative of the total of 143 veterans in the trait measured—scholastic aptitude. It should be remembered, however, that there are only two chances out of a hundred that the difference between the group studied and the total General College population is the result of chance factors.

Further description of the sample is provided by the distribution of courses taken as shown in Table 2.

TABLE 2  
*Distribution of 248 Courses Taken by the Sample*

<i>Course</i>	<i>Per cent of courses</i>
Written Expression .....	9.5
Social Science . . . . .	29.0
Natural Science . . . . .	11.0
Literature . . . . .	6.5
Mathematics . . . . .	4.5
Art and Music . . . . .	5.5
Vocational Studies . . . . .	8.0
Personal Studies . . . . .	22.0
Speech . . . . .	4.0
	100.0

It will be noted from Table 2 that slightly more than half of the courses taken are in the areas covered by the GED Tests.

<sup>3</sup> The mean and standard deviations of the distribution of raw scores for the sample on this test were 62.55 and 18.60, respectively.

<sup>4</sup> The latest data available for a typical General College population were for the 1941 freshman class.

The honor-point ratios of the sample ranged from 3.0 to -0.43, with a mean and standard deviation of 1.36 and 0.81, respectively. (Three honor points are given for a mark of A, with other values ranging to -1.0 for a mark of F.) Since in the previous studies the *Ohio State University Psychological Examination*, form 21, has proven to be one of the best single predictors of scholastic success (5, 6), scores on this test were correlated with honor-point ratios of the sample to furnish a basis for comparison of this study to previous prediction studies. Standard scores on the GED Tests were used in this study since it can be reasonably assumed that the distribution from which these scores were computed was normal (7).

### Findings

The total GED Tests scores were found to correlate 0.72 with first term honor-point ratios of the sample.

TABLE 3

*Relationship between Honor-Point Ratio and the GED Tests and the Ohio Test*

	<i>N</i>	Honor point ratio ( <i>r</i> )	$\sigma_r$	5% confidence interval
GED Total . . . .	56	0.72	0.064	0.60 to 0.85
GED I . . . . .	58	0.51	0.097	0.32 to 0.70
GED II . . . . .	59	0.60	0.083	0.44 to 0.76
GED III . . . . .	59	0.53	0.091	0.37 to 0.73
GED IV . . . . .	58	0.56	0.090	0.38 to 0.73
Ohio,* #21 . . . .	48	0.62	0.090	0.44 to 0.80

\* The Ohio Test scores correlated with Total GED Test scores 0.71,  $\sigma_r$  of 0.075, and 5% confidence interval of 0.56 to 0.86 ( $N=45$ ).

This is higher than the coefficients between the honor-point ratio and any individual GED Test or the Ohio Test. The GED Test II (Social Science) was found to be the best individual GED Test as a predictor of scholastic success ( $r=0.60$ ). By inspection the GED Tests seem to have much in common with the content of courses in the General College. This is particularly true for the GED Test II (Social Science). Table 3 shows the relationship between honor-point ratios, and GED Tests scores and the Ohio Test scores for the sample, and Table 4 shows the distribution of standard scores obtained by the sample on the GED Tests.

TABLE 4

*Distribution of Standard Scores Obtained by the Sample on the GED Tests*

	<i>N</i>	Mean	S.D.	Range
GED Total . . . . .	56	218.62	25.27	172 to 285
GED I . . . . .	58	48.28	6.80	38 to 65
GED II . . . . .	59	57.82	8.63	39 to 79
GED III . . . . .	59	57.99	8.76	38 to 78
GED IV . . . . .	58	53.40	7.72	38 to 69

*Comparison of This Study to the Yale Study*

Since this study and the Yale study (2) are somewhat similar, it is desirable to know how well the two agree. It is possible to compare these studies by using Kelley's formula (4, 222) for predicting a correlation coefficient for a different range of talent group.<sup>5</sup>

$\frac{\sigma_1}{\sigma_2} = \frac{\sqrt{1-r_{22}}}{\sqrt{1-r_{11}}}$ , where 1 signifies the short range of talent, and 2 signifies the long range of talent.

Using this formula to predict the results of this study from the results of the Yale study, it was found that for the total GED Test scores and GED Test II (Social Science) scores the predicted  $r$  and the obtained  $r$  were essentially the same. The widest discrepancy between the predicted and the obtained  $r$

TABLE 5

*Prediction of Correlation Coefficients for This Study from the Findings of the Yale Study*

	Predicted $r$	Obtained $r$
GED Total . . . . .	0.71	0.72
GED I . . . . .	0.59	0.51
GED II . . . . .	0.60	0.60
GED III . . . . .	0.75	0.55
GED IV . . . . .	0.63	0.56

<sup>5</sup> It is questionable whether this formula may be used validly for this purpose. Kelley describes its use in predicting reliability coefficients for different ranges of talents; also, it assumes that the two samples being compared are from the same total population. It is a special case of a more general formula. In its use here it controls only one variable, leaving the other variable (honor-point ratio) uncontrolled. It is offered here only as a possible comparison technique and no claim is made for its validity since the basic assumption may not have been met and only one variable has been controlled.



was for GED Test III (Natural Science). These results are shown in Table 5. Considering the possible differences in the samples used and in the nature of the curricula, there is a substantial degree of agreement between this study and the Yale study.

### *Conclusions*

The following conclusions seem to be warranted on the basis of this study and the comparison of it with other prediction studies:

1. The total GED Tests score is better than any of the individual GED Tests as a predictor of scholastic success for the sample used.
2. The GED Test II (Social Science) is the best individual GED Test as a predictor of scholastic success for the sample used.
3. Of the measures studied and reported to this time, the total GED Tests score shows promise of being one of the best single predictors of scholastic success.

Although these tests were not designed as predictors of scholastic success, they appear to serve this capacity quite well. The above conclusions are highly tentative, since they are based on a small sample, and need further study and verification on larger and different populations before they can be accepted with a high degree of confidence.

### REFERENCES

1. Bradley, Mary E. "A Study of the Validity of the Armed Forces Institute Tests of General Educational Development in the Field of Social Studies." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 265-268.
2. Crawford, Albert B. and Burnham, Paul S. "Trial at Yale University of the Armed Forces Institute General Educational Development Tests." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, IV (1944), 261-270.
3. Dyer, Henry S. "Evidence on the Validity of the Armed Forces Institute Tests of General Educational Development (College Level)." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, V (1945), 321-333.
4. Kelley, Truman L. *Statistical Method*. New York: Macmillan Company, 1923.

5. Monroe, Walter S. (editor). *Encyclopedia of Educational Research*. New York: Macmillan Company, 1942.
  6. Segel, David. *Prediction of Success in College*. A Handbook for Administrators and Investigators Concerned with the Problem of College Admittance or Guidance of College Students. Washington: Government Printing Office, 1934.
  7. *United States Armed Forces Institute, Tests of General Educational Development (College Level), Examiner's Manual*. New York: American Council on Education, 1944.
-

# AN ANALYSIS OF FOUR ORALLY ADMINISTERED MEASURES OF ADJUSTMENT

WILLIAM D. ALTUS  
Santa Barbara College

AND

HUGH M. BELL  
Chico State College

DURING World War II all soldiers who were functionally illiterate at the time of their induction into the Army were sent to a Special Training Center, where they attained an approximate fourth-grade level of literacy within twelve weeks or were discharged for inaptness. Half of the trainee's time while at a Center was spent on the drill field; half was spent in the classroom. A number of variables conditioned a trainee's chances of graduation. His literacy on arrival at the Center, his adjustment to the Army, and his mental capacity were all important to his survival.

In a previous study by Altus and Bell (2), the total number of maladjusted responses on four orally administered tests of adjustment was shown to have a biserial correlation of  $.453 \pm .028$  with the disposition of the trainee, i.e., whether graduating from the Center and retained in the Army or failing to graduate and being discharged for inaptness. The most valid of all tests employed by the Personnel Consultants' Section in its association with the criterion of disposition is the group literacy test used for placing men in one of four instructional levels. The biserial validity of this placement test with the disposition of the trainee is  $.794 \pm .017$ . Adjustment correlated with this placement test  $.108 \pm .028$ , the coefficient of correlation being the Pearson product-moment linear type. The Pearson  $r$  for adjustment and mental ability, as determined by the total score on four verbal subtests of the *Wechsler Mental Ability Scale*,

Form B, was  $.152 \pm .028$ . The more maladjusted trainee was slightly less literate and not quite so bright as the well-adjusted. The biserial correlation of the Wechsler with the disposition of the trainee was  $.521 \pm .013$ .

It is obvious from the data in the foregoing paragraph that a trainee's adjustment is second in importance only to his initial literacy, so far as predicting his type of disposition is concerned. The Wechsler measure of mental ability was of some value in the team of three tests here discussed but its inter-correlation of  $.43 \pm .035$  with the highly valid placement test made it less useful than adjustment for prediction purposes. Initial literacy, adjustment and intelligence were important in that order to the Personnel Consultant's forecast of the trainee's chances of graduation.

The four tests of adjustment were Army Adjustment, Hypochondria, Paranoia, and "Concentrated Bell." The derivation of these tests has been previously discussed (2). The individual items composing the several tests appear later in this article. In the form in which they appear in this article, the four tests were used for some months prior to mid-December, 1943, after which time the tests were combined into a single measure of adjustment described elsewhere (1). The test-retest reliability of the fifteen questions purporting to measure hypochondria was .71. For the 25-point measure of Army Adjustment the test-retest reliability was .81 and for the "Concentrated Bell" the  $r$  was .86. The reliability of Paranoia was not checked. When all four tests were used as a single scale, the odd-even, stepped-up reliability proved to be .957 for 100 cases. The 87 items composing the total scale thus appear to be as reliable as many well-known tests of general aptitude.

### *I. The Validity of the Four Adjustment Tests*

The biserial correlation of the four adjustment tests with the criterion of trainee disposition was  $.453 \pm .028$ . This figure unfortunately gives only an index of over-all relationship and does not tell much about the sensitivity of the total scale at its extremes. In order to check the differential validity of varying levels of tested maladjustment, the total maladjusted

responses of 507 trainees were tabulated and then divided into quintiles. Quintile V consists of the most maladjusted one-fifth of the trainees; Quintile IV is the next most maladjusted one-fifth. Quintile I, as would be inferred, consists of the best-adjusted one-fifth of the trainees. Only 9 per cent of Quintile I received such discharges; 11 per cent of Quintile II; 14 per cent of Quintile III; 25 per cent of Quintile IV; 38 per cent of Quintile V. It appears, therefore, that the combined adjustment test is sensitive throughout the whole range but is most sensitive at the maladjusted end of the distribution. The least-well-adjusted one-fifth of the trainees (Quintile V) received over four times as many discharges as did the best-adjusted one-fifth of the trainees (Quintile I).

A further test was made of the sensitivity of extremes in the maladjusted score by comparing the percentage of inaptness discharges for the best-adjusted one-tenth of the trainees with the percentage received by the most poorly adjusted one-tenth. Only 6 per cent of the best-adjusted 10 per cent of all trainees received inaptness discharges while 42 per cent of the most poorly adjusted 10 per cent were so disposed of; in other words, seven times as many inaptness discharges were to be found among the most poorly adjusted one-tenth. Still another comparison was made, this time for the highest and lowest 5 per cent in adjustment scores. Of the best-adjusted 5 per cent of the trainees, only 4 per cent received inaptness discharges; of the least well-adjusted 5 per cent, 59 per cent were discharged for inaptness. The ratio for this last comparison is almost fifteen to one. The tremendous range in adjustment scores found among the extreme 5 per cent at both ends of the scale should also be mentioned. The best-adjusted 5 per cent had from zero to four maladjusted responses out of eighty-seven possible; the least well-adjusted 5 per cent had a minimum of fifty maladjusted responses. The sharp cleavage in score between these groups is no doubt mainly due to the rigorous selection of the adjustment items through successive item analyses.

The data presented in the two preceding paragraphs are based on 507 cases. In another study of 300 cases, it was found

that there was a tetrachoric correlation of .50 between adjustment and disposition of the men at the Training Center. This  $r$  compares favorably with the biserial correlation of .453 previously reported. At least three factors (1) operated to reduce the correlation of adjustment with the disposition of the trainee. One was the race or language group to which he belonged. Adjustment is quite important to the disposition of the non-English, Spanish-speaking trainee, to the Indian and to the Negro; adjustment is of considerably less importance to the bilingual Mexican and to the old-line, native-born American White (1). The screening of inductees by the psychiatrist at the time the men were inducted into the Army also markedly reduced the range of possible maladjustment. All of the trainees who arrived at the Center had been accepted as being sufficiently well adjusted for military service. The third factor limiting the effectiveness of the adjustment test was the literacy which the trainee had on his arrival at the Center. For instance, in this study of 300 cases, 176 out of 183 trainees above a certain literacy level eventually graduated; the seven who were discharged were, however, maladjusted. Within the same group of 300, only three trainees out of fifty-nine who were below the accepted minimum level of literacy eventually graduated; all three were well adjusted. Adjustment is of most significance for those who have an approximately even chance of graduating if they apply themselves with normal diligence, but as can be seen it is also of considerable import even among the extremes of relative literacy.

This discussion has shown that adjustment is of marked significance in its association with the type of disposition received by the trainee. The validity of the adjustment test is a function of the whole range of testable adjustment, not of limited segments of it. Since there is little correlation between tested adjustment and either intelligence or literacy, it is obvious that adjustment, *per se*, is the operating factor. It may now be of interest to consider the separate items of the four tests, their respective item analyses and, finally, their tetrachoric correlation with the criterion of disposition of the trainee.

II. *Individual Adjustment Items*

*Army adjustment.*<sup>1</sup>—The twenty-five items in the Army Adjustment test appear in Table 1. The first column shows the discrimination of the items for the upper and lower quartiles in the total adjustment score. This analysis is the third and

TABLE 1  
*The Validity of the Army Adjustment Items*

Item analysis*		Validity†	Item
Q <sub>4</sub>	Q <sub>1</sub>		
20	7	-.11	1. Do you like this camp?
17	2	.20	2. Do you like the food here at this camp?
7	1	-.03	3. Do your clothes fit you?
5	0	.16	4. Are your shoes comfortable?
16	4	-.08	5. Do you think they ought to mix men up the way they do here?‡
16	5	-.07	6. Have you been used to associating with the type of men we have here?
19	5	.07	7. Do you feel lonely here?
16	4	.02	8. Are the men too noisy in your barracks?
15	2	.00	9. Are the latrines too dirty?
20	0	.14	10. Do you get very homesick?
19	3	.14	11. Does it bother you to sleep and wash with so many men?
9	3	-.10	12. Do you think soldiers get enough passes?
16	1	.25	13. Did your draft board give you a raw deal?
22	6	.00	14. Do you think you could do more for your country as a civilian than you can as a soldier?
18	2	-.16	15. Did you expect to have a tough time in the Army?
19	2	.20	16. Did you get a good physical examination at the Induction Station?
22	2	.32	17. Do you think you are strong and healthy enough to be a soldier?
2	1	.00	18. Have any of the non-coms been mean to you?
23	1	.37	19. Do you think you are a nervous person?
18	2	.28	20. Does it bother you to stand and wait in a group?
7	1	.07	21. Did you have enough time to straighten out your home affairs before you came into the Army?
20	7	.33	22. Do you worry much about your dependents?
21	9	.44	23. Does it bother you that you can't read and write very well?
15	1	.07	24. Do you think that this school here will help you?
18	7	.02	25. Do you think that you will be happy in the Army?

\* Analysis is based on upper (Q<sub>4</sub>) and lower (Q<sub>1</sub>) quartiles in maladjusted score for 100 trainees.

† Tetrachoric coefficients inferred from computing diagrams devised by Chesire *et al.* (4). All the coefficients in the following three tables were computed in the same manner. Population consists of 200 cases, 100 of whom were discharged for inaptness and 100 of whom were graduated and shipped.

‡ Question used because Chinese, Filipinos, Indians, Mexicans, and old-line American Whites were quartered together.

<sup>1</sup> Not an official Army test    Used experimentally at the Training Center.

final one through which this test was passed. It will be noted that, in general, the discrimination is relatively good. The most discriminating item is the tenth one; 30 per cent of the upper quartile answered this item in a maladjusted fashion while none whatever of the lower quartile did. Most of the remaining items markedly differentiated the two quartile groups but with an item such as the eighteenth, "Have any of the non-coms been mean to you?" too few answered the item in a maladjusted fashion to justify its inclusion. This is the only item among the twenty-five which has meager discriminating power between the quartile groups.

The validity coefficients for each item are somewhat too high. For all of the tests in the four tables which follow, an equal number of inapt and graduate trainees is used—a figure which markedly overstates the number of discharges and also distorts the coefficients. The distortion, empirically checked, is about .05 for an  $r$  of .50. Since the distortion is constant for all items and since the relative, not the absolute, validities were desired, the slight exaggeration of the size of the  $r$ 's does not vitiate any of the discussion which follows.

A marked range in the validating coefficients will be noted. The most valid one, with an  $r$  of .44, is the one, "Does it bother you that you can't read and write very well?" Not only invalid but also inversely related to the disposition of the trainee is the question, "Did you expect to have a tough time in the Army?" The  $r$  for this latter question is  $-.16$ . The five best items include the one previously cited and, in descending order of validity, items 19, 22, 17 and 20. Five items in addition to item 15 have an inverse validity. They are items 1, 3, 5, 6, and 12. It will be noticed that the most valid items are highly personal and appear also to have an emotionally charged character. The invalid items are generally of a highly specific nature ("Do your clothes fit you?") and are not so personally probing.

There is some reason to believe that a trainee's first reaction to specific features of military establishment (barring the type of food served; cf. item 2) is not of great importance to his eventual disposition. It is rather interesting that the items



measuring dislike for certain racial and language groups (items 5 and 6) apparently are not related to the final outcome of his training. The first item shows that a trainee's initial feeling toward the Training Center has no association with his ultimate disposition. Highly specific complaints about external conditions of Army life appear to be more variable and less significant than the personal and emotionally charged ones, such as 23, 19, 22, 17, and 20.

The dispositional validities and the item analyses of the separate questions were given rank order and the resulting series were correlated by the rank-difference method. The resultant coefficient was .39. It is probable that the restriction in range brought about by the successive item analyses is partly responsible for the comparatively low coefficient. The fact that there is a relation, though low, between the internal validation of the test by item analysis and by an outside criterion is evidence, for this one test at least, of the correctness of the techniques employed to construct an adjustment test. The final proof of the validity of any individual item is, however, its co-variation with the criterion; if it is not correlated or is inversely correlated with the criterion, the item proves to be worthless for the specific purpose intended. The best item from the standpoint of validity, question 23, is among the six poorest from the standpoint of item analysis; conversely, question 15, which has a negative validity of  $-.16$ , stands up well in the item analysis. But the individual correlational validity of each item must be the final judgment, no matter how good or how poor the question may appear from item analysis.

*Paranoia.*—The same data for the Paranoia test are shown in Table 2 as were shown for the Army Adjustment test in Table 1. The Paranoia items analyzed somewhat better than did the Army Adjustment items, especially for those in the lowest quartile of the maladjusted score, since most of the latter group had no maladjusted responses whatever on the Paranoia test. The Paranoia scale of the *Minnesota Multiphasic Personality Inventory* (5) was the original basis for the present test. The original questions were re-phrased and shortened after the first analysis, and some new items were added. This

revised test was then further item-analyzed; more questions were discarded and more added. The final four questions in the test were intended to be used clinically; they were believed to be too strong, even for Army illiterates, for them to have

TABLE 2  
*The Validity of the Paranoia Items*

Item analysis*		Validity†	Item
$Q_1$	$Q_4$		
30	0	.23	1. Have people refused to help you when you were down and out?
29	3	.24	2. Have you had more hard luck than most people?
42	6	.18	3. Have others had it easier than you have had?
31	0	.14	4. Would you have got along better if people hadn't stood in your way?
30	0	.35	5. Do you often feel that the whole world is against you?
34	0	.27	6. Have you had an awful lot of trouble?
35	0	.26	7. Do you feel lonely most of the time?
31	0	.25	8. Do you take things harder than most people?
16	0	.20	9. Do you sometimes feel that you are being followed?
22	0	.44	10. Do people pick on you a lot?
40	4	.03	11. Do you often feel that people do not understand you?
30	0	.29	12. Have people played dirty tricks on you when they had no reason to?
24	0	.12	13. Has your hard luck been due to other people's meanness?
14	0	.20	14. Were your parents too strict with you?
21	1	.28	15. Would you fight anyone who got in your way?
25	2	.05	16. Do people talk about you behind your back?
28	1	.13	17. Have you had hard luck all your life?
15	0	.09	18. Have the bosses you've had worked against you?
11	0	.20	19. Do a lot of people have it in for you?
11	0	.20	20. Have you had a lot of run-ins with the cops?
17	3	.31	21. Does the Devil take hold of you at times?
8	0	.04‡	22. Do you know who is the cause of your troubles?
15	0	.07‡	23. Is there someone working against you behind your back?
6	1	.05‡	24. Has anyone ever tried to poison you?
6	0	.05‡	25. Does one of your enemies have control of your mind and make you do things you don't want to do?

\* Analysis is based on 200 cases, 100 white and 100 colored. Each quartile has 50 cases represented in it.

† Tetrachoric coefficients based on 200 cases, 100 being discharged for inaptness and 100 graduating from the Center.

‡ There were insufficient maladjusted responses to these four items to justify credence in the obtained  $r$ 's. These questions were meant for clinical use—that is, to help determine gross abnormality.

validity in the same sense that the other questions might have. It will be noted that the data in Table 2 bear out this inference.

The Paranoia items are more valid than the Army Adjustment items; no items with negative validity appear in the

former test while they do in the latter. The tenth item is the most valid in Table 2, the coefficient being .44. The next best items in descending order of validity are questions 5, 21, 12, and 15. Questions 10, 5, and 12 appear to have much the same objective reference—that of feeling other people were deliberately unkind and malevolent—though their inner meaning to the trainee is obviously different because all are of unequal validity. Item 21, "The Devil takes hold of me at times," should perhaps be explained. The trainee, in answering the question in a maladjusted fashion, appeared to take it in a metaphorical sense, that is, some force so strong that it seemed to come from external sources bent him toward a line of action he knew to be wrong—and this force is here personified for him by the "Devil." The item shows very excellent validity. One of the five best items, item 15, is one of open aggression—"I'll fight anyone who gets in my way." It is the only question among the twenty-five that taps overt expression of paranoiac reference.

The least valid questions are items 11, 13, 16, 17, and 18, the last four questions not being considered as part of the regular scale. It may be that these five items do have paranoiac reference, but it appears from their relative lack of validity that if they are truly measures of paranoiac tendencies they must represent a part of the continuum which is characterized by a depressive inertness which does not eventuate in overt expression. Two of these items, 13 and 17, have reference to one's hard luck; three others, 2, 3, and 6, also touch upon the same question. The five "luck" items do not have psychic equality, however, for the validity coefficients range from .12 to .27. One may perhaps infer from this finding that it is sometimes desirable in a test of adjustment to include a number of variations on the same theme, for validity is thereby increased.

A glance at Table 1 will show that item 7, which has to do with feeling lonely, has an  $r$  of .07; item 7 of the Paranoia test (cf. Table 2) also touches on loneliness but its coefficient is .26. A slight re-phrasing, shifting the connotative value of a question, sometimes markedly changes its meaning and its validity. There is also the possibility that the ordinal position of an oral

adjustment item may affect its validity. There is a perfect correspondence for the four tests between their respective total validities and the order in which they were administered. Army Adjustment was the first test administered to the trainee; it also has the least total validity. The "Bell" test is the last in the series; it has the most validity of all. Whether this finding is purely adventitious is not known.

The relation between the validity of the individual Paranoia items and their rank in goodness from the standpoint of item

TABLE 3  
*The Validity of the Hypochondria Items*

Item analysis*		Validity†	Item
$Q_1$	$Q_2$		
22	0	.22	1. Do you feel tired when you get up in the morning?
18	0	.20	2. Do you feel well most of the time?
22	1	.32	3. Do you get dizzy often?
21	2	.48	4. Does your heart or your lungs bother you?
15	0	.27	5. Do you get short of breath easily?
23	3	.32	6. Are you in good health most of the time?
17	0	.47	7. Do you have trouble sleeping at night?
19	2	.25	8. Do your eyes often bother you?
19	3	.37	9. Does your hip or back bother you?
19	1	.35	10. Do you often have headaches?
13	0	.36	11. Does your stomach bother you very much?
20	2	.10	12. Do your legs often hurt you?
21	4	.28	13. Do your feet hurt you when you walk a lot?
7	0	.20	14. Have you taken much medicine in your life?
9	1	.16	15. Have you had to go to many doctors for treatment?

\* Based on 100 cases, twenty-five in each quartile.

† Tetrachoric correlations based on 200 cases, 100 graduates and 100 discharges for inaptness.

analysis is  $-.01$ , the last four items not being used in the correlation. Why there is no relationship for these two variables in this test when some relationship is found for the other three is not apparent.

*Hypochondria.*—The Hypochondria test was also derived in part from the *Minnesota Multiphasic Personality Inventory*. The Hypochondria test also went through three item analyses. The original items were shortened, deleted, and changed after each analysis and some new items were added. It will be seen in Table 3 that the individual validities and item analyses are

quite good. The item analyses range from .22 and 0, for question 1, to .7 and 0 for question 14. The validities range from .10 to .48, all being positive. Two items have coefficients approaching .50; one of them has reference to insomnia and the other pulmonary and cardiac pain. The questions relating to hip-back pain and to stomach upset are quite good as are the items referring to frequent headaches and to dizziness. Except for the insomnia question, all of these items point to the very common type of psychosomatic symptom, used as a refuge by the psychically lame.

Two of the items, 2 and 6, have reference to one's general state of health. Though apparently of equal value so far as objective meaning goes, they are not psychically equivalent. The  $r$  for the first question is .20; for the second it is .32. Again one sees the marked effect that wording has upon the validity of some questions.

The rank-difference correlation of the item analysis validity with the dispositional validity of the questions is .17 for the Hypochondria test. The relatively small difference among the questions in their item analyses may be the attenuating factor.

*Bell.*—The twenty-two questions comprising the "Concentrated Bell" are the most valid items from Bell's *Adjustment Inventory, Student Form*, which would be within the ken of adult male illiterates after the items were simplified by rephrasing and shortening. The original item analysis of these questions (3) for high-school and college students is given in the first column of Table 4. In the second column is given the item analysis for the Army illiterates. Very marked discrimination among quartile groups is shown for both analyses. The validity coefficients for the separate items are also quite high, bearing out the previous study (2) which showed the "Concentrated Bell" to have the greatest amount of dispositional validity of all of the four adjustment tests.

The most valid item is number 20, a question dealing with insomnia. Its coefficient of validity, .49, is almost exactly the same as that for the insomnia question in the Hypochondria test, .47. The different manner of asking these two questions is not reflected in their respective validities, which are practi-

TABLE 4  
*The Validity of the Concentrated Bell Items*

Item analysis*	Item analysis†		Validity‡		Item
1	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>1</sub>		
8	2	89	7	.37	1. Do you often feel sad and blue?
8	2	70	10	.19	2. Do you often feel lonesome even when you're with people?
2	13	87	8	.34	3. Do you feel very tired toward the close of the day?
6	0	27	0	.26	4. Do you cry very often?
13	2	64	6	.20	5. Are you troubled with shyness?
17	3	71	3	.34	6. Do you have trouble starting to talk with someone you've just met?
18	3	68	18	.14	7. Do you make friends with women easily?
18	13	58	25	.12	8. Are you often sorry for the things you do?
15	5	70	36	.43	9. Do your eyes hurt you very much?
3	0	80	4	.28	10. Do you often feel just miserable?
32	2	80	9	.45	11. Are your feelings easily hurt?
17	0	31	4	.20	12. Do you make friends easily?
19	7	65	5	.31	13. Does (Did) either of your parents get angry easily?
32	0	59	5	.32	14. Does it bother you to be criticized—that is, have people say you haven't done something right?
33	0	65	7	.32	15. Do you stay off by yourself when you go to a party?
53	2	77	20	.05	16. Do you have to watch your health carefully?
72	7	75	17	.27	17. Is (Was) either of your parents very nervous?
93	2	57	3	.42	18. Do you find it hard to talk to a stranger?
75	0	27	5	.38	19. Do you feel that neither of your parents understands (understood) you?
70	3	84	1	.49	20. Do you have trouble sleeping at night because you lie awake thinking about things?
88	0	62	2	.27	21. Does it bother you to come into a room where a lot of people are sitting around talking together?
87	9	58	4	.16	22. Do you daydream a good deal?§

\* An item analysis in terms of percentages of maladjusted responses found among the most maladjusted (in terms of adjustment score) fifteen per cent, Q<sub>1</sub>, and the best-adjusted fifteen per cent, Q<sub>2</sub>. Cf. Bell (3).

† An item analysis, in terms of actual numbers, of 444 cases, 111 of the highest 15 per cent in maladjusted (according to twenty-two items here recorded) responses, Q<sub>1</sub>, and 111 of the best adjusted 25 per cent, Q<sub>2</sub>, in terms of maladjusted score on the 22 items here given. All cases in this analysis were trainees.

‡ Tetrachoric coefficients based on 200 cases, 100 of whom were graduates, 100 of whom were discharged for inaptness.

§ This item frequently had to be translated in a roundabout, circumlocutional manner, varying with the vocabulary of the given trainee.

cally synonymous. Item 11, "Are your feelings easily hurt?" is second in validity, the  $r$  being .45; third in validity is the "eye-strain" question, number 9, with a coefficient of .43. The "feeling" item obviously refers to social sensitivity while the other item belongs to a hypochondria group. Item 18, diffi-

culty in talking with strangers, has an  $r$  of .42; almost the same question is asked in item 6 but the coefficient for this latter item is .34. The fifth question in order of validity is the one relating to parents in item 19; the tetrachoric coefficient is .38. Emotional problems in the home apparently left a deep impress upon the functioning of these illiterates, for two other parent-subject questions have quite good validity in addition to item 19. These two are question 13, relative to the ease with which the parents became angry, and question 17, relative to the nervousness of the parents. Among the six most valid of the 22 items is only one depression question, number one in the test, "Do you often feel sad and blue?" There is only one other true depression item in the test and that is number 10.

Items relating to social sensitivity and to emotional scars from early parent-child relationships appear to be the most significant ones among the Bell items, though depression and hypochondria are also of some importance. What makes the Bell items of considerable interest, in addition to their high validity, is the fact that they were also the best items for a population strikingly different from that involving the Special Training Center trainee. The original standardization group was young, literate, and of a relatively superior economic level. Also the original group consisted of girls as well as boys. It may be of parenthetical interest to note here that 104 college girls, tested in 1943, proved to be significantly more maladjusted in mean scores on the 22-point "Concentrated Bell" than did the trainee group. It seems that these questions come much closer to being "universal" than analogous aptitude questions. Aptitude questions valid for Army illiterates would be too easy to possess validity for, say, college girls. Yet within the adjustment continuum, the Bell questions are valid for both quite disparate groups. Adjustment questions appear able to cut across sex, aptitude, and economic boundaries in a manner seldom found for analogous aptitude questions.

In spite of the quite narrow range of discrimination power among the 22 Bell items, there is a rank-difference correlation of .35 between item analysis validity and dispositional validity. This coefficient confirms the similar correlation found for Army

Adjustment and, to a lesser degree, for Hypochondria. Why no relationship was found for the Paranoia test is unknown. In any event, it is clear that both types of validation may be used with profit, though if a testing device is tailored for a specific purpose the item correlation method must be the final criterion of validity.

### *Summary*

1. The biserial correlation of the total maladjusted responses on four adjustment tests with the disposition of illiterate soldiers sent to an Army Special Training Center was  $.453 \pm .028$  for 507 cases. For 300 other trainees, the tetrachoric correlation of the same variables was  $.50$ .

2. When used with a highly valid group literacy test for predicting the disposition of the trainee, adjustment was shown to be preferable to aptitude as measured by an individually administered test.

3. The combined maladjusted score on all four adjustment tests was shown to be valid throughout the tested range. Over four times as many of the trainees in the highest 20 per cent of the maladjusted score received inaptness discharges as did the best-adjusted 20 per cent. The rate of inaptness discharge for the best-adjusted five per cent was four in 100; for the least well-adjusted five per cent, the rate was fifty-nine in 100, the ratio between the two extreme groups being almost one to fifteen.

4. The highly specific items relating to a trainee's initial impression of an Army camp were shown to be of negligible value in their association with his disposition, i.e., whether graduated from the Center or discharged as inapt. Questions relating to personal adjustment with a rather heavy emotional charge were, on the other hand, relatively valid.

5. The individual items in the Paranoia test were generally quite valid. Certain of the items which showed good item analysis discrimination had little dispositional validity. Validity was seen to be partly a function of the manner of phrasing a question, for the same thought expressed in five different ways showed considerable differences in validity.

6. Insomnia and the common psychosomatic complaints



concerning headaches, stomach trouble, pulmonary and cardiac pain were among the most valid items for the Hypochondria test.

7. The "Concentrated Bell" items were the most valid of all. The items in this test relating to social sensitivity and scarifying parent-subject experiences proved to be the most valid items in terms of the disposition of the trainee.

8. For three of the four tests there was a positive relation between item analysis validity and dispositional validity, ranging from a rank-difference correlation of .15 for Hypochondria to .39 for Army Adjustment. The lowness of these relations may be attributed to the attenuating effect of the narrow range of item analysis validity.

9. It is suggested that dimensions of adjustment are more inclusive than those of aptitude, for while certain items, such as the "Concentrated Bell," are valid for Army illiterates as well as for college girls, aptitude questions valid for one group would generally possess little validity for the other.

10. It is shown that orally administered adjustment tests, if properly selected by item validation, do not necessarily suffer from lack of reliability. The odd-even, stepped-up reliability of the four tests, combined, is .957.

11. The successful experiment with orally administered adjustment tests for Army illiterates suggests that similar techniques might be of value elsewhere, perhaps in industry and in education.

#### REFERENCES

1. Altus, W. D. "The Adjustment of Army Illiterates." To be published.
2. Altus, W. D. and Bell, H. M. "The Validity of Certain Measures of Maladjustments in an Army Special Training Center." *Psychological Bulletin*, XLII (1945), 98-103.
3. Bell, H. M. *Theory and Practice of Personal Counseling*. Stanford University, California: Stanford University Press, 1935.
4. Chesire, L., Saffir, M. and Thurstone, L. L. *Computing Diagrams for the Tetrachoric Correlation Coefficient*. Chicago: University of Chicago Bookstore, 1933.
5. Hathaway, S. R. and McKinley, C. *The Minnesota Multiphasic Personality Inventory*. Minneapolis: University of Minnesota Press, 1943.



# THE ROLE OF THE PSYCHOLOGIST IN THE COUNSELING PROGRAM OF THE VET- ERANS ADMINISTRATION<sup>1</sup>

GWENDOLEN SCHNEIDLER

Advisement and Guidance Service, Veterans Administration, Washington, D. C.

THE Vocational Rehabilitation and Education Office of the Veterans Administration has the legal responsibility for the vocational rehabilitation of disabled veterans who take training under Public Law 16, as amended, and for the education and training of veterans who pursue courses under Public Law 346, as amended.

As the first step in the vocational rehabilitation of disabled veterans, a counseling service is required. Veterans who apply and who are eligible for benefits of the Servicemen's Readjustment Act may receive counseling service upon request. The Office of the Assistant Administrator for Vocational Rehabilitation and Education of the Veterans Administration has through the Advisement and Guidance Service attempted to meet this legal responsibility in a number of ways:

First, it has designed and established policies and plans for a counseling program in Veterans Administration field offices. This has been accomplished mainly through the efforts of the Director of the Advisement and Guidance Service, Dr. Ira D. Scott, and his staff. These policies and procedures have been presented in systematic form in the *Manual of Advisement and Guidance*, published by the Government Printing Office.

The next step which the Advisement and Guidance Service took to fulfill its legal obligations was to establish an operational organization in Veterans Administration field offices. Advisement and Guidance Sections were established in sixty-

---

<sup>1</sup> Paper presented at the American Psychological Association Convention, September 5, 1946, University of Pennsylvania, Philadelphia

three Veterans Administration regional offices. In the beginning of the program these sections were to constitute the offices in the state to which veterans could come for counseling services. After having had some initial experience and having selected and trained a staff, they were to decentralize, to establish guidance centers at educational institutions, to train advisers for these centers and to supervise them and administer the counseling program for the region.

The guidance centers, of which there are now over 350, are the actual operating units where the counseling of most veterans is accomplished. The Veterans Administration contracts with the educational institution to provide certain types of services in return for a fee for each veteran counseled. Whenever possible the institution furnishes, in addition to space, personnel who function as counselors and psychometrists.

At each guidance center there is a minimum Veterans Administration professional staff consisting of one vocational adviser, who is designated Chief of the Veterans Administration guidance center, and one training officer. The Chief of the guidance center explains to each veteran the educational benefits to which he is entitled and reviews the final records to ascertain whether the employment objective finally selected by the veteran with the assistance of the counselor seems appropriate. The training officer then follows through and assists the veteran in locating and enrolling in an appropriate educational institution or training establishment.

The director appointed by the educational institution at which the Veterans Administration guidance center is located is responsible for his staff of counselors and psychometrists and for their selection, training and supervision, to make certain that the highest professional standards are maintained in interviewing, completing the records, testing and counseling.

In addition to the guidance centers, counseling sections are being established in many Veterans Administration, Navy and Army hospitals in order that a veteran, whose physical condition warrants it, may start his educational and vocational planning with the assistance of a vocational adviser or may even begin training in line with his long-range objectives while he is still in the hospital.

The third major activity of the Advisement and Guidance Service has been that of establishing Civil Service standards for the selection of professional personnel. The extent to which this has been successful will soon be evident after the Civil Service has administered an examination to the vocational advisers who have been hired. The central office has encouraged the development of this examination since it was realized that many factors, such as a lack of qualified personnel, especially during the war years, and the system of local rather than centralized control of the selection of personnel may have contributed to the lowering of standards in some areas.

A fourth activity of the Advisement and Guidance Service has been the initiation of a training program for advisers. In addition to conferences held in various parts of the country and attended by most of the vocational advisers, it is planned now to establish three types of training programs:

1. Orientation training—This program will be to acquaint new employees with Veterans Administration regulations and procedures related to counseling.
2. In-service training—This will include staff conferences and institutes.
3. Professional improvement—It is expected that this will comprise a combination of institutional training and on-the-job training under a program which will allow new trainee-employees to work half-time in trainee positions as vocational advisers and to take formal training in psychology at the graduate level, the cost of instruction to be paid by the Veterans Administration.

A fifth area of activity of the Advisement and Guidance Service has been that of supervision. This has not been as extensive as desirable, but since the organization of the thirteen branch offices, the main function of which is supervision, it is expected that a much more continuous supervision program will be possible, with supervisors in the field at least half the time.

The Advisement and Guidance Service has recognized that educational and vocational counseling also often involves personal counseling and therefore it has initiated a program in-

volving the selection, training, and assignment of personal counselors to all regional offices and to as many guidance centers as possible. The qualifications for the personal counselor call for the completion of certain courses related to clinical and differential psychology plus about three years of qualifying clinical or teaching experience involving personal counseling.

In addition to these types of activities through which the Advisement and Guidance Service has attempted to fulfill its functions, it has also been furnishing field offices with technical tools and services, in such areas as:

1. Tests and measurement.
2. Occupational information.
3. Educational and vocational counseling.

These are some of the ways in which the Veterans Administration, through its Advisement and Guidance Service, has attempted to meet its legal obligations. It has planned and put into operation a counseling program involving the use of psychological methods and techniques and has tried to secure the services of qualified personnel as vocational advisers, psychometrists, personal counselors, and supervisors for a program which is perhaps more ambitious than any of its kind ever put into operation in the field of education or psychology.

These facts should be of interest as well as of some concern to psychologists. Since psychological methods are being utilized on such a wide scale, psychologists should desire to see that they are being used by persons who are well trained and who will bring credit to their use. The profession of psychology has no legal responsibility for the counseling of veterans as does the Veterans Administration, but psychologists should have a feeling of professional as well as of social responsibility. It would seem that they should recognize the character of this emergency, as they did during the war years, and should make whatever contributions they can.

This raises the pertinent question of what psychologists as individuals and as members of a professional organization can do to help the Veterans Administration meet the present emergency in its program for the vocational rehabilitation and counseling of veterans.

The most direct way, but certainly not the only way or even the best way in many cases, to deal with psychological problems within the Veterans Administration is to become a full-time employee of this governmental agency. A number of psychologists have already entered the employ of the Veterans Administration. In the central office Advisement and Guidance Service, for example, one third of the professional staff are members of the American Psychological Association. The proportion is less for field offices, although there is scattered throughout the country an increasing number of psychologically trained persons, although not so many as the central office considers desirable.

There are still some positions open in guidance centers, regional offices, branch offices and in the central office at the P-2 to P-6 classifications with salaries ranging from \$3397 to \$7102. There is need for personal counselors at the P-4 level which carries a beginning salary of \$4902. It is suggested that any qualified persons who are interested should watch for the announcement of the Civil Service examination covering all professional positions in the Advisement and Guidance program of the Veterans Administration and should apply to take it, because all future appointments to permanent positions will be made from registers which will be established as a result of this examination.<sup>1</sup>

There are some psychologists who are not interested in employment in large governmental agencies where the top administrative personnel are not psychologists and where the profession is not represented at the highest policy-making levels, but where psychologists might be considered as technicians serving the organization. Until the profession of psychology is represented at top levels, at least through advisory committees, many good applied psychologists will prefer to remain in the more "protected" educational institutions where they are not required to apply psychology themselves except by teaching students how psychological methods should be applied in business, industry, and government.

---

<sup>1</sup> Examination now scheduled for February 6, 1947.

But if the training of the present generation of graduate students does not include some orientation and attention to their social and professional responsibilities in addition to their own professional development, it will be unfortunate for the agencies requiring their services, for veterans and others needing competent psychological service, and particularly for the profession itself, because the job must and will be done, if not by personnel who are best qualified by professional training, then by others who are not as well qualified. Unless a program such as that of the Veterans Administration is able to attract a sufficient number of well-qualified and trained personnel, there may be serious repercussions for the profession. The general public will recognize that psychological methods are being used, but they may sometimes be dissatisfied with the way these methods are used and with the results obtained.

One way for psychologists to take an active part in the Veterans Administration counseling program and at the same time to retain their positions in academic institutions, is to be associated with one of the Veterans Administration guidance centers which are now located on over 350 campuses. A number of psychologists are now serving either part or full time as directors of the educational personnel at such guidance centers or as counselors or psychometrists.

An original purpose in establishing contracts with educational institutions was to enhance both the quantity and the quality of the Veterans Administration counseling service by capitalizing on the well-qualified personnel who were already on the staffs of psychology departments and who would not need extensive training to counsel veterans. It was recognized that many psychologists could not be lured away from their established academic positions, nor did it seem desirable to do so. But it is apparent now that the Veterans Administration was overly optimistic in thinking such institutions would have available on their staffs sufficient numbers of persons with experience in educational, vocational, and personal counseling to serve veterans at these centers.

In the first place, colleges and universities lost many such personnel to the armed services and, because of the heavy post-



war veteran enrollment and the needs for instructional personnel, they are finding it necessary to place those who returned in full-time teaching positions, where they are lecturing to veterans in large classes rather than counseling with them individually.

Furthermore, many colleges and universities had had no adequate program of personnel services for their own students and consequently had had little similar experience from which to profit. What counseling had been done was academic rather than vocational and personal and was limited to college students rather than adapted to the needs of the wide range of the general population of adults.

Instead, therefore, of having the best talent of the country available as counselors in Veterans Administration guidance centers, new and sometimes poorly qualified personnel have been hired by educational institutions and placed on their non-academic staffs without any assurances of tenure and with low salaries. This situation is not characteristic of all, however. Some institutions are selecting well-qualified professional counselors and paying them adequately. But, in general, colleges and universities have hired personnel for these positions in competition with the higher Civil Service salaries. As a result, many such counselors, instead of being better qualified than Civil Service employees, are often not as well qualified. Dr. J. G. Darley has found, for example, in a sample survey of fifty-five guidance centers in May and October of 1945, that only 22 per cent of the counselors and psychometrists on the guidance center staffs of these institutions were members of the American Psychological Association or of the American College Personnel Association.

There are great differences among the Veterans Administration guidance centers in the services being rendered to veterans and most of these differences can be attributed to differences in quality of leadership and of assigned personnel. The Veterans Administration supplies one vocational adviser and sets certain minimum standards against which the work results produced by the institutional personnel are expected to be checked. But, otherwise, the professional counseling is accomplished for the

most part by the institution's personnel. In some centers counselors are well qualified in training and experience, they are trained in the correct procedures, are adequately supervised, and are well paid and motivated to perform at a high professional level. In others, the institution either does not have an understanding or it does not have a high regard for the quality of professional services which should be rendered. It takes little administrative interest in the center, regards it as another contender for the limited space available and fails to assign to it a well-qualified, full-time director and qualified professional counselors.

Psychologists at institutions having Veterans Administration guidance centers should, it would seem, take an active interest, if not an active part, in the work of the guidance centers.

A number of psychologists will soon be given an opportunity to serve as consultants to the Veterans Administration. Plans have been completed to invite selected psychologists to serve on panels of consultants. A list has been prepared and agreed upon by representatives of the American Psychological Association and the Veterans Administration for twelve psychological consultants for the central office. Each of the thirteen branch offices will also have a panel of three to six such consultants selected in a similar manner with the concurrence of the Deputy Administrators. These persons will be invited to render consultation services when called upon by personnel of the Veterans Administration Division of Clinical Psychology and Advisement and Guidance Service. They will be paid according to the amount of service they render.

There will be additional opportunities, by means of contracts with the Veterans Administration, for other psychologists to assist in the training of Veterans Administration and institutional personnel serving as counselors and psychometrists. For example, in accordance with a contract made by the University of Chicago and the Veterans Administration, a series of six-week courses in personal adjustment counseling is being conducted for personal counselors. Other universities will be asked to conduct similar courses and their psychologists will have an opportunity to function in a similar manner.

In addition, plans are being developed to make available to Veterans Administration personnel needing to increase their professional competence, other special short courses and institutes at colleges and universities in such fields as tests and measurements, counseling techniques, occupational information, statistical and research methods, and problems of counseling the physically disabled.

In addition to these suggestions for individual psychologists who wish to serve the Veterans Administration, there are other services which could be rendered by psychologists as members of the American Psychological Association. For example, committees representing the A.P.A. might be selected to keep informed of the Veterans Administration programs which utilize psychological tools and methods, to advise administrative officials of the Veterans Administration regarding any special problems of interest to the profession, to suggest procedures not being used which are considered desirable or to make other suggestions to improve psychological services to veterans.

There are undoubtedly still other ways in which psychologists might assist the Veterans Administration with its vocational rehabilitation and counseling program. The Veterans Administration is facing an emergency situation in the rehabilitation of veterans to civilian life which is just as real, if not just as great, as the emergency which the Army and Navy faced in the classification, training, assignment, and adjustment of personnel for military and naval duties. The Advisement and Guidance Service is attempting to meet the Veterans Administration's legal responsibility, and it is believed that psychologists have a social and professional responsibility as well as an opportunity to make a contribution to this program in the present emergency.

# COMMUNITY RELATIONSHIPS IN OUT-OF-SCHOOL COUNSELING

WARREN K. LAYTON

Detroit Public Schools

## *Background*

In the early 1930's certain problems of Detroit's youth were in the forefront of community consideration. One concrete expression of this concern was a study entitled "What of Youth Today?" directed by a cooperative staff representing several agencies; two of the principal workers were Dr. Rachel Stutzman of the Merrill-Palmer School and Mr. Ray Johns of the Detroit Young Men's Christian Association. Two outstanding conclusions in this study were that there existed in the community (1) a lack of coordination of youth services and (2) markedly inadequate facilities for the counseling of out-of-school youth.

The first step in correcting these conditions was the formation of the Detroit Council for Youth Service, a voluntary organization of 25 persons (representing agencies), whose purpose was to secure the cooperation of all youth-serving agencies to the end that their programs would be more effective. The Board of Education as the chief youth agency participated in the founding of the Council and provided office space and telephone service, and the Superintendent of Schools has always served as Chairman of the Council. The Staff salaries and all other costs are paid by funds provided by a private foundation. The Executive Secretary is Miss Claire M. Sanders. Almost simultaneously with the establishment of the Council the National Youth Administration requested a local advisory committee; a special committee of the Council took over this function and continued in this capacity until the end of the NYA program in 1943.

Since the Council for Youth Service is a representative group (education, public services, social services, management, labor, citizens-at-large), and since its purpose is to coordinate and to assist youth agencies, it has never been thought feasible for the Council to set up any new services directly serving youth. However, when the NYA began to operate its project programs for unemployed out-of-school youth, it became apparent at once in the assignment of youth to projects that a counseling service of some type was essential. Accordingly, the Detroit Council for Youth Service, with NYA cooperation and funds, organized in 1936 a joint agency for counseling NYA youth known as the Junior Consultation Service, appropriating the title of the well-known New York agency.

The Detroit Junior Consultation Service was housed in a downtown office building in which the Council for Youth Service had its headquarters and in which the Department of Guidance and Placement of the Detroit Public Schools also maintained offices. The last-named agency supervises the work of about 150 school counselors and has operated the principal placement referral agency for individuals under twenty-one years of age, directly to employers, since 1922. Inevitably the new Junior Consultation Service and the regular placement service of the school system found much in common.

In 1937 a new Michigan State Employment Service, established under the Wagner-Puyser Act, began operations. This was, and is, by far the best public employment agency Michigan has had. Its Assistant Director, now its Director, Mr. Edward L. Cushman, an able and enterprising as well as a highly trained executive, is community-minded and cooperative. Since MSSES was responsible for the certification of NYA youth, a close relationship was built up between it and the Junior Consultation Service.

The need for a larger coordinated counseling program for youth and younger adults now became increasingly apparent. Through the efforts of Miss Sanders, Mr. Cushman, and others interested, a new agency known as the Detroit Counseling Service was established in 1938, with its own broadly representative Executive Board. The principal sponsoring agencies were

the Employment Service, the NYA, the Detroit Public Schools, and the Council for Youth Service. The Schools provided space, telephone service and other overhead expenses, the part-time services of the Director of Guidance and Placement as Director of the new agency, and the full-time services of Miss Milma Wickstrom as Supervising Counselor. The Employment Service provided psychologists and clerical workers, and the NYA an Intake Counselor and the staff counselors. The staff numbered 16 persons. A full counseling service was thus made available to all persons from sixteen to twenty-five years of age. Unhappily the Detroit Counseling Service has not been able to continue on this basis indefinitely, owing to the withdrawal of Employment Service support coincident with federalization of all the state employment services in 1942, and to the discontinuance of NYA in 1943.

### *Present Status*

The idea of out-of-school counseling had been firmly established in the community by this time, however, and despite the demands of the war effort upon all agencies, including the public schools, the school officials felt that the Detroit Counseling Service should be continued. The volume of work had declined somewhat as more and more young persons entered the armed services and the war industries; many also took the places of others in civilian employment. Since 1943, the Detroit Public Schools have operated the Detroit Counseling Service alone on the same pattern as before. The staff at present consists of two staff counselors, a psychologist especially competent as a vocational analyst, and two clerical workers, all on a full-time basis; the part-time services on a supplemental basis of any of the five staff counselors of the Department of Guidance and Placement, of the Director, the receptionist, and an additional clerk.

### *Types of Counseling Problems*

While environmental influences such as the war itself, war-time employment, return from war, and strikes, have influenced our young people from time to time, we find that individuals

seeking counseling present mainly the same problems. Miss Wickstrom classifies our clients as follows:<sup>1</sup>

1. Those who have made no plan, or only a vague or partial plan, who need help in discovering the occupation which they would like to enter, and who need information about how best to prepare for it.
2. Those who have made a plan unsuited to their abilities. For example, the boy who wishes to be a tool and die man because he has heard of the advantages, yet who has received his poorest marks in related subjects and who, on tests, reveals little aptitude in this field.
3. Those who have no realization of what their chosen occupations involve, like the girl who wants to be a health education teacher because she likes basketball and other sports.
4. Those who because of economic necessity have had an unwise decision forced upon them, like the boy who for three years was a helper on a delivery truck and who on the *American Council Achievement Tests in Mathematics, English, and Science* obtained scores in the 99th and 100th percentiles, and would like to prepare himself for engineering.
5. Those who are in conflict with their families over a choice of plans, like the girl who wants to become a nurse, but whose parents insist upon her becoming a stenographer because of probable immediate employment.
6. Those who have lost interest in school, but who through assistance in making an occupational plan begin to realize the importance of further training and would like to make plans to get it. The problem here is increased of course, if a long period of time has elapsed since leaving school.
7. Those who are drifters and need to be encouraged to find immediate employment which will stimulate them in establishing better habits, like the boy who hitch-

---

<sup>1</sup> *The American Child*. New York: The National Child Labor Committee, January, 1946.

hiked and wandered about the country for a year, sleeping in parking lots and all night movies, and who now would really like to settle down, but does not know where to begin.

8. Those who have special personal problems and who need help in making adjustments before success in any occupation may be possible, like the girl who because of an unhappy home situation and many conflicts, adopts a defensive attitude and feels that all of the adult world is her enemy.
9. Those who have made good plans, but who desire confirmation of their choice.
10. Those who have severe handicaps, mental, physical, or emotional.

### *Testing Programs*

Aptitude testing has been an important phase of the work of the Detroit Counseling Service from its inception. No one standard battery is used regularly, and in some instances no tests are used. As an illustration of the testing service, in January, 1946, some of the tests used were as follows:

Achievement: *Cooperative Examinations in Chemistry, English, General Culture, Literary Comprehension, Social Problems, and Vocabulary;*

Clerical Aptitude: *Minnesota Vocational Test for Clerical Workers;*

Intelligence: *American Council Psychological Examination; Otis Self-Administering Test of Mental Ability; Pressey Senior Classification Test;*

Vocational Interest: *Kuder Preference Record;*

Manual Dexterity: *Purdue Peg Board;*

Mechanical Aptitude: *Minnesota Paper Form Board;*

Personality: *Bernreuter Personality Inventory; The Personal Audit.*

### *Results Achieved*

An effort is made each month to obtain follow-up information on what has happened to our clients. In October, 1945, we learned that twenty-seven of our clients received employ-



ment, thirty-three had returned to high school, twenty-six were enrolled in some special training program, seven had entered a college or university, four had received medical care, and ten had undertaken needed recreation.

### *Current Trends*

At present, returning service people constitute a large portion of our clientele. Their median age is between twenty and twenty-three years, and their problems are largely educational and vocational in character. Prior to V-J Day the service people who sought our help presented problems of personal adjustment, very largely. Prior to the war period, our clients were for the most part in the age group from eighteen to twenty-four and their problems centered about occupational planning and job seeking. In 1943 and 1944 the clients were younger, and personal and educational problems were in the ascendant. Throughout the period the Counseling Service has been in operation it has been noted that there is a tendency for self-referrals to increase markedly.

### *Interpretations*

The following seem to us to be important factors in our experience with this out-of-school counseling program:

1. The entire community has contributed at one time or another in various ways to the project, and has a permanent interest in it through the Executive Board.

2. The assumption by the public school authorities of the professional and financial responsibility for supporting in part the planning group known as the Detroit Council for Youth Service, and first partially and then wholly the Detroit Counseling Service, is significant as a recognition of the educational authority as an agency for service to *all* of the community's youth.

# ESTIMATION OF TEST ITEM DIFFICULTY BY AVERAGING HIGHEST AND LOWEST QUARTER PERFORMANCE COMPARED WITH TOTAL POPULATION COUNT<sup>1</sup>

LIEUTENANT COMMANDER DANIEL D. FEDER  
United States Naval Reserve

A SEVENTY (70) item final achievement examination was administered to one hundred and forty-four graduates of an officer technical school. Conventional item analysis was made, obtaining  $R_{Hta}$  with a total score for each item by counting successes among the highest and the lowest 27 per cent of the population, converting to percentages, and then entering the Flanagan tables. Difficulty values of each item were calculated by averaging the two percentages thus obtained. One item was deleted because of technical errors which made its exclusion from the test necessary. The distribution of total scores yielded an essentially symmetrical normal curve.

It was desired to determine whether the above-described method of determining item difficulty yielded values significantly different from those which might be obtained by a full count of the performance of the entire population. Therefore, the next step was to obtain the average per cent of difficulty for each item using the entire group of one hundred and forty-four (144) papers. The difference between the two difficulty values was then computed for each item. These differences were as follows:

<i>Differences between the Percentage Difficulty Values</i>	<i>Number</i>
- 8	1
- 7	0

<sup>1</sup> The opinions and assertions contained in this paper are those of the writer and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

-6	0
-5	4
-4	3
-3	7
-2	8
-1	5
0	14
1	10
2	6
3	5
4	0
5	3
6	1
7	0
8	2
Total	<u>69</u>

It will be noted that the distribution is an approximately symmetrical normal distribution for chance errors. The mode and median are zero (0).

Neither magnitude nor sign of the differences bore direct relationship either to the difficulty value or discriminatory value of the items. For example, the two plus 8 differences were for items having  $R_{its}$  of .47 and .48. The minus 8 difference item had an  $R_{its}$  of .16. Check of all items for which the difference was  $\pm 4$  or more yielded no possibility of a generalization concerning possible relationship to the bi-serial coefficients.

It is concluded, therefore, that the method of computing difficulty values by averaging the percentage of success in the highest and the lowest "quarters" yields substantially the same results as would be obtained by dealing with the entire range, for a symmetrical total score distribution.

# APPLICATION OF INFORMAL PROJECTIVE METHODS IN THE COUNSELING INTERVIEW<sup>1</sup>

HAROLD B. PEPINSKY<sup>2</sup>

Michigan State College

THE purpose of this paper is to show that informal projective materials, forming part of the clinical counselor's office furnishings, can be used effectively in the counseling interview. The hypothesis is advanced that responses to such materials need not be standardized since they form a part of the dynamic interview process and vary from client to client.

It is interesting to note that the majority of reports involving projective techniques have centered on the study of individuals whose behavior is abnormally deviant in terms of our cultural norms and hence properly within the province of the clinical psychologist (7, 9). On the other hand, the clinical counselor, who is concerned with the diagnosis and treatment of minor functional maladjustments (4, 10), has apparently made little use of projective methods.

There appear to be three major reasons for this:

1. The clinical counselor's training and experience rarely permit him to delve extensively into emotional problems.
2. His lack of medical affiliation forces him to be extremely sensitive to public opinion concerning the nature of his work.
3. As a balance against public opinion and lack of clinical background, he has been forced to rely heavily upon psychometric test data and other relatively objective sources of information.

---

<sup>1</sup> This article represents a modification and expansion of a paper by the writer (3) read before the Division of Guidance and Personnel Psychologists, A.P.A., at Philadelphia, September 4, 1946.

<sup>2</sup> The writer is indebted to J. B. Holland and Pauline N. Pepinsky for their helpful criticisms.

As a counterbalance, however, the nondirective counseling movement, led by Rogers (6) and by Roethlisberger and Dickson (5) has influenced the clinical counselor to focus upon what happens *during* the counseling interview and especially upon the "attitudes and behavior" of the *client* (6, p. 3).

Thorne (8), though praising the contributions of nondirective psychotherapy, has criticized this system of reflecting feelings back to the client. One of Thorne's criticisms is that the use of the method may result in a superficial analysis of the client's problems, hence a treatment of symptoms rather than causes. Bordin (1) and Pepinsky (2) have urged that directive and nondirective methods be validated against the kinds of problems presented by the client.

Pending definitive experimental evidence on the adequacy of either approach, the writer recommends the use of informal projective techniques as a part of the counseling interview. The method does seem to be an effective compromise, especially in the diagnosis and treatment of emotional conflict problems:

1. The client is helped to release pent-up, emotionally charged feelings.
2. The counselor is able to probe for causes of the client's problems without arousing resentment or defensiveness in the client.
3. The responsibility for analyzing and solving his problems can remain with the client.

The writer came upon the idea of using projective techniques informally in the counseling interview in the fall of 1944. He was employed at that time as Assistant Director of the University of Kansas Guidance Bureau, and one of his major duties was that of counseling university students and other adult clients relative to their educational, vocational, and personal problems.

A 17-year-old freshman girl, whom psychometric tests had shown to be a highly intelligent, well-adjusted individual, had become increasingly "nervous" (to use her expression) and had received mid-term grades far below her predicted achievement level. A verbatim report on a dramatic incident in the interview follows:

Counselor: You would like to relax. Would it help you to look at my picture here? (Points to picture on wall.)

Client (after a brief pause): "I'm looking at it, but I don't like it.

Counselor: Can you tell me more about that?

Client (clenching her fists): The branches of that tree—they seem to be clutching for something they never reach—just like me. (Bursts into tears.)

The client appeared to feel better as the result of this release. It was relatively easy from then on for the counselor to help her verbalize the conflict she had felt in coming from a small rural community, in which she had played an important role, to the more cosmopolitan and impersonal college community, in which her role was relatively unimportant.

The picture itself was a lithograph measuring 11 by 16 inches, depicting a bleak landscape with no people visible in it. During the year 1944-45, the counselor made verbatim records of responses to it by fifty additional clients, all of whom were university students, the majority freshmen, in the age range sixteen to forty-three years.

Protocols of responses plus supplementary case-record materials suggest the lithograph's value in the counseling interview. There was the case of a nineteen-year-old junior in the School of Pharmacy who claimed to be undecided in her vocational choice. She was an attractive girl, always well dressed, a member of a sorority, and a campus leader.

The counselor directed the client's attention to his lithograph during the course of a counseling interview. He was surprised at her concentration on the small details in the picture, many of which he had been unaware of.

Client (after description): Well, that's all I can think of, I guess.

Counselor: Can you tell me how you feel about the picture?

Client: Well (pause), it makes me want to go in and help the people who live there.

Counselor: You want to go in and help the people?

Client: Why yes—I've always wanted to help other people . . . maybe because I have so many problems to work out myself.

Whereupon, the client launched into a lengthy discussion of her personal problems: Her parents had been divorced and

both laid heavy claims to her affections. She, however, felt most at home with an aunt in a neighboring city. She finally decided to align herself with the aunt, and seemed greatly relieved. Her vocational choice remained unchanged.

Soon afterward, a girl with a similar problem came to the counselor for help. The parents had been divorced and the father was about to marry a woman to whom the girl was greatly attracted. The mother appeared to be highly neurotic and accused the father of stealing the girl's affections. Some of the girl's responses to the lithograph are given here:

First Interview: "Oh golly, that picture looks peaceful—I'd like to go up in those mountains and take a long walk by myself. I'd feel so much better just getting away from everything."

Subsequent Interview: "Gosh, that picture always makes me feel so good. I know I can always find peace in those mountains—away from the turmoil of the city."

This client subsequently decided to seek her happiness in the father's new home.

Many additional illustrations might be given. For example, a young man who was having academic difficulties primarily because of intellectual limitations, told stories about life in the community which he envisaged in the lithograph, stories in which he took an active part in the simple "folksy" life of the villagers. A returned serviceman, discharged as an N. P. case, saw peace in the village but dreaded the ominous clouds overhead. An extremely intelligent girl who was reminded unpleasantly of the Kansas dustbowl came to see the conflict between her recent emancipation and her conservative home background, which she rebelled against but dreaded leaving.

The counselor found each client's response so uniquely relevant to his own problems that standardization could not be attempted without the protocols of numerous additional clients and without the confirming reports of other professional counselors. Furthermore, it was not possible for him to differentiate between the use of the lithograph in diagnosis and in therapy.

The lithograph did appear to serve three major functions:

1. The client was permitted to feel at ease since the stimu-

lus-situation was seemingly external. Thus free responses were more readily forthcoming, tension appeared to be released through the expression of emotionally toned ideas, and the client appeared better able to develop self-insight as the result of looking outside of himself at his problems.

2. The counselor felt that he was able, without direct questioning, to clarify his diagnosis by uncovering hidden motives.
3. The responsibility for analysis and the solution of his problems remained with the client.

Subsequent empirical research involving the use of other media, such as a game of anagrams and free drawings, has been equally promising. The counselor has been able to employ these materials informally in the counseling interview, without making the client suspicious or hostile to what he might otherwise regard as an intrusion into his private world.

In conclusion, informal projective techniques, when utilized by the professionally competent counselor, may well prove a real economy in time and effort over other interview approaches, especially in the treatment of emotional conflicts.

#### REFERENCES

1. Bordin, E. S. "Diagnosis in Counseling and Psychotherapy." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 169-184.
2. Pepinsky, H. B. "The Reliability of a Set of Problem Categories for Use in the Clinical Counseling of College Students." Ph.D. Thesis, University of Minnesota, 1946.
3. Pepinsky, H. B. "Informal Projective Techniques in Client-Centered Counseling." *The American Psychologist*, I (1946), 250. (Abstract.)
4. Pepinsky, H. B. "Therapeutic Counseling." C. E. Erickson, ed., *A Basic Text for Guidance Workers*. New York: Prentice-Hall, in press.
5. Roethlisberger, F. J. and Dickson, W. J. "The Interviewing Method." *Management and the Worker*. Cambridge: Harvard University Press, 1943, pp. 270-291.
6. Rogers, C. R. *Counseling and Psychotherapy*. Boston: Houghton-Mifflin Company, 1942, 445 pp.
7. Sargent, Helen. "Projective Methods: Their Origins, Theory, and Application in Personality Research." *Psychological Bulletin*, XLII (1945), 257-293.



8. Thorne, F. C. "A Critique of Nondirective Methods of Psychotherapy." *Journal of Abnormal and Social Psychology*, XXXIX (1944), 459-470.
9. White, R. W. "Interpretation of Imaginative Productions." J. McV Hunt, ed., *Personality and the Behavior Disorders* (2 vols.). New York: Ronald Press, 1944. I, 214-251.
10. Williamson, E. G. *How to Counsel Students: A Manual of Techniques for Clinical Counselors*. New York: McGraw-Hill, 1939, 562 pp.

## COUNSELING AND THE MINNESOTA POINT OF VIEW

E. G. WILLIAMSON  
University of Minnesota

It has long been my personal opinion that concepts, theories, systems of thought, and personal service programs are best understood in terms of two sets of data usually not available to those not participating in the early formulation of these concepts and programs. I refer first to the *institutional setting* in which a concept arises and which modifies the concept in many subtle ways. For example, it is my hunch that the so-called non-directive system of thought about counseling has been profoundly influenced by the fact that it arose largely out of experience with the younger-age clients who, in our culture, are habitually and traditionally treated in family relationships in a relatively more indirect and subtle manner than are adults. Other influencing factors, in addition to the age, developmental status, and sex of the client, include the kind of institution in which the original work was done, such as a university or a community child guidance clinic or industrial personnel department. A second major source of information about a system of counseling has to do with personal and professional experiences of the originator and leader of the movement. For example, as I shall point out, the so-called Minnesota counseling system of thought was profoundly influenced by the personal experiences of its originator, D. G. Paterson, who essentially was attempting to adapt and adopt from industrial psychology and army personnel experiences certain techniques and points of view relevant to the development of a method of vocational guidance worthy to be described by the word, "valid." In a real sense, the personal experiences of the originator contributed certain peculiar characteristics to the movement itself.

Believing that this point of view about the evaluation and

understanding of systems of counseling may have some value in the current discussion of counseling methods, and particularly those that appear to be in conflict with each other, I shall review briefly some of the evaluative phases and peculiar characteristics of the so-called Minnesota point of view. I use the phrase "so-called Minnesota point of view" because as far as I have been able to discover in twenty years of work at Minnesota, there is no unified and one-person-dominated point of view, but rather a series of points of view, which, in certain respects, are closer to each other than they are to some other points of view, such as the non-directive point of view. I shall review these common elements or common characteristics which, as far as I can understand my colleagues, we hold in common. I point out as a footnote that just as "individual differences" is a fundamental doctrine among Minnesota psychologists, just so should the reader keep in mind that the many "uncommon" and herein unmentioned, divergent points of view are also necessary to a full understanding of the Minnesota point of view. The right, indeed the necessity, of testing these divergent views experientially and experimentally is even more important a part of our doctrine, as Paterson has always maintained.

Let me describe briefly the institutional setting in which the Minnesota program developed. The early beginnings go back to the year 1914 when the late Dean J. B. Johnston became Dean of the Liberal Arts College and began a search for valid and dependable methods of identifying, prior to registration and enrollment, those students whose scholastic potentialities were such as to make it very unlikely or improbable that they would succeed in college, and conversely, to identify those whose potentialities were high and who were, therefore, likely to succeed in college. The dean made numerous investigations applying scientific methods to a study of college administration problems, a notable contribution in itself. His works are best summarized in his publications.<sup>1</sup>

---

<sup>1</sup> Johnston, J. B. *Scholarship and Democracy*. New York: D. Appleton-Century Co., Inc., 1937. Pp. 113.

Johnston, J. B. *Who Should Go To College?* Minneapolis: The University of Minnesota Press, 1930. Pp. 22.

Johnston speaks of his own work in the following terms which are quite relevant today in view of the new and current influx of students into colleges, seeking as did earlier students, to insure their high-level occupational futures by means of a college education:

For several years before 1914 College faculties were concerned over the number of incompetent students coming into college. The writer entered administrative work with the hope that a remedy for this could be found by discovering the evidences of aptitude for college work and creating an interest in the guidance of youth into the kinds of work that were most suitable for them. . . .

. . . It is now evident that the college growth that began in 1914 was a mass movement expressing an undefined and only half intelligent desire of the people for more of the benefits of education.<sup>2</sup>

Elsewhere Johnston restates his social and educational philosophy including the point that each student must be assisted to discover that type of education and that vocation which are congruent with his aptitudes and interests. In this manner, misuse of talents can be avoided, and society as well as the individual can thus be benefited. Johnston assigned to education and guidance societal functions of major importance and the direction and force of this point of view are still in evidence in Minnesota.

### *Testing and Personnel Work*

In the present brief review of the Minnesota counseling program, I shall select only seven significant characteristics of Minnesota programs in order to high-light just what it is we have been trying to do during the past several decades following the early leadership of Johnston. The first characteristic is the emphasis on testing as one of many features of a well-rounded personnel program. Johnston's early concept of developing reliable and valid and objective methods of identifying potentialities of students has been a dominant frame of reference for at least one large part of the program although by no means for the entire program. This has usually led to the publication, for non-Minnesotans, of test prediction studies,

---

<sup>2</sup> Johnston, J. B., *op cit.*, p. 12.

and the impression thus has been erroneously created that testing is all that is done in the counseling program at Minnesota. For example, Rogers says, in reference to an article by the Bixlers:

This article points up the fact that even in a setting where students and counselors alike have regarded tests as the center of all counseling, a client-centered approach brings about a very different orientation on the part of the client, and a very different use of tests.<sup>3</sup>

Since it was our decision in 1944 to "try-out" the so-called non-directive point of view in counseling by "importing" a psychologist trained with Rogers, I may be permitted to comment about Rogers' footnote remark without implying any criticisms or lack of confidence in the effective work of Bixler. We did this not because we believed tests were "the center of all counseling," but rather because it has long been the fundamental strength of Minnesota to try out new ideas and techniques, regardless of the source. For us the validity test of a counseling technique has been "what results does it produce and under what conditions," not "does it fit in with tests."

The results of this importation, mutually gratifying to Bixler and to us, is another story to be told elsewhere. But such a pragmatic approach to the validity testing of counseling techniques does not in any way indicate agreement with Rogers' characterization of the Minnesota program as exclusively or even largely test-centered. The program has always been broader than testing and the use of tests, and the following documentation will serve to buttress this rejoinder.

In the April, 1925, Vol. III, issue of the *Journal of Personnel Research*, Paterson described the student personnel services of the University of Minnesota under that title. In this very early period of the Minnesota program, the very broad scope of the activities of Paterson and his staff of personnel workers is indicated by the following list of center headings in that article:

- |                                      |                                     |
|--------------------------------------|-------------------------------------|
| 1. Testing intelligence tests        | 3. Orientation courses for freshmen |
| 2. Committee on educational guidance | 4. Personnel advice                 |

---

<sup>3</sup> Rogers, Carl R. "Psychometric Tests and Client-Centered Counseling." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 142, footnote 2.

- |   |                               |
|---|-------------------------------|
| 5. Preparation of the advisers            | 10. Personnel research        |
| 6. The personnel record cards             | 11. The marking system        |
| 7. Case histories                         | 12. Efficiency of instruction |
| 8. Vocational consultants                 | 13. Student mortality         |
| 9. Gifted students and probation students | 14. Activities                |
|   | 15. Future progress           |

In April, 1928, the supplement of the *Educational Record*, Vol. 9, pp. 1-40, contains a reported survey by Paterson's committee on the student personnel program at Minnesota. This report is a review of progress made up to 1927, and ends with thirty recommendations having to do with aspects of the broad personnel program. It should be pointed out that relatively few of these recommendations and of the topics discussed in the long report are restricted to tests and testing phases of counseling. The report begins with a discussion on organizing student personnel services quoting extensively from Scott and Clothier's *Personnel Management*, with regard to the necessity of decentralization of personnel services. That is, the principle was enunciated in these early days that personnel work should be the function of every staff member of the University and not merely of experts. Parenthetically, it may be pointed out that this concept of personnel work as a function of every member of the institution is quite different from the conception which sometimes arises from professional experience in a child guidance clinic.

The report goes on to discuss ten projects which are listed below:

1. The Necessity for Educational Vocational Guidance in the High School. Dean J. B. Johnston.
2. Personnel Information to be Secured from the Students at the Time of Admission. D. G. Paterson.
3. Measuring Aptitude for College and for Specific Curriculum Courses. D. G. Paterson.
4. Curriculum Provisions for Individual Differences, Dean J. B. Johnston.
5. Educational and Vocational Guidance for Undergraduates. D. G. Paterson.
6. Improving our Examining and Grading Systems as a Necessary Step in Basing Guidance in Part upon Achievement in College Courses. Clare M. Brown.
7. The extent to which Extra-Curricular Activities Facilitate or Inhibit Adjustment to College Life. Vernon Williams.

8. The extent to which Financial Aid and Opportunities for Employment are Facilitated in the Adjustment of Freshmen. Vernon Williams.

9. Student Health and Mental Hygiene. H. S. Diehl.

10. Coordination of Personnel Activities. D. G. Paterson.

Two later surveys indicate the continuing broad concept of student personnel services which has been maintained at Minnesota during these past decades. I refer to the survey I published with R. Sarbin entitled, "Student Personnel Work in the University of Minnesota,"<sup>4</sup> consisting of 113 pages of description of the many personnel activities in all parts of this institution. A later survey on counseling and its many manifestations and ramifications was prepared by Dr. Harold Pepinsky, published (mimeograph) in 1945 under the title, "Counseling at the University of Minnesota." This latest report describes the counseling done by faculty members, by advisers, and by various specialized individuals, including psychologists in the Student Counseling Bureau.

My point needs no further elaboration. At Minnesota, from the very beginning, the program of counseling has been conceived as an integral part of a student personnel program and not as a separate isolated service of experts and technically trained persons alone. Testing and counseling are two of a number of personnel services. To those psychologists who are not familiar with a student personnel program and its many ramifications, I would suggest reading of the above references, and of Lloyd-Jones and Smith's comprehensive book, *A Program of Student Personnel Service for Higher Education*.<sup>5</sup>

By way of further orientation, it should be pointed out that during the past eight years, we have been engaged in working out applications of psychological counseling principles, techniques, and methods, to several types of personnel services which are not usually performed in colleges by psychologically trained counselors. For example, counseling methods are being applied in a new counseling program of financial assistance of students.<sup>6</sup> In addition, counseling methods have been

<sup>4</sup> Minneapolis. Burgess Publishing Company, 1940.

<sup>5</sup> New York: McGraw-Hill Book Company, 1938.

<sup>6</sup> See No. 7, Series VI, *Student Personnel Work*, on "Financial Aid for College Students" for a preliminary outline of this type of counseling. In part this brochure is based upon George Risty's experiences in financial counseling.

worked out and applied to the handling of disciplinary cases among college students and an extensive description of this radically new counseling program at the college level will be published soon. A new program is under way involving the development of psychological methods of counseling foreign students enrolled in the University. We also are working on the application of counseling methods, techniques, and principles in the utilization of organized activities as personalized means of helping individual students learn social and emotional adjustment through participation.<sup>7</sup> This is in effect an attempt to develop a new integration at the college adolescent level of three different methodologies: individualized counseling, group work, and group therapy. Lastly, we are all involved in the perfecting of appropriate psychological counseling and group work in the dormitories and fraternities by psychologically trained counselors employed full time or as part-time graduate students in psychology or educational psychology.

### *Improvement of Vocational Guidance*

My second characteristic of the Minnesota program is the use of various devices, such as high-school scholarship, to identify students who will probably experience great difficulty in meeting the institution's academic and scholastic standards. This was an attempt to develop valid methods of vocational and educational guidance to substitute for the then current methods which the late President Coffman described as "often the merest piffle." A cursory reading of the literature of guidance in the 1910 to 1925 period and a comparison with the current literature will highlight the progress made in this field by applied psychologists. Measurement and testing, then, were brought into the field of vocational guidance as a means of helping the individual student avoid misjudging his own aptitudes and interests and were further introduced into the counselor's training to help the counselor avoid similar pitfalls.

---

<sup>7</sup> Williamson, E. G. "Student Activities: An Integral Part of Student Personnel Work." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VII (1944), 3-12. Further descriptions of the early stages of this program are in preparation by Theron Johnson and Barbara Clark.



At Minnesota, educational guidance based upon, but not restricted to, tested aptitudes made significant gains in the early identification of potentially failing students and the use of dependable counseling methods to help them to choose substitute or alternative or more appropriate goals. It was Dean Johnston's conception that such direct assistance to students before they actually begin failing in their studies would not only help the institution itself through releasing its energies and resources to do a better job with the potentially able student, but it would also prevent much of the emotional frustration and maladjustment so prevalent among the low-ability students. The reader will see that this point of view differs from the concept of the role of tests described by Rogers.<sup>8</sup> At Minnesota, it has always been felt that scholastic mortality was a waste of individual and social resources and, therefore, early prevention through intelligent and friendly counseling was necessary and desirable.

### *Fallibility of Judgments*

A third major characteristic of the Minnesota program is Paterson's insistence from the early beginning that human judgments, subject as they are to all types of errors, should be refined and improved in validity, as well as reliability, through the use of objective devices of one sort or another, including aptitude tests and achievement tests. I have always felt that the modern training of personnel workers is seriously deficient with respect to this very significant point of orientation. As far as I know, personnel workers of today, and particularly those going into the college field, are seldom if ever introduced to the literature summarized first in Hollingsworth's "Judging Human Character." As a result of this deficiency, many counselors are not aware of the pitfalls and errors of judgment-making, even those of the counselor himself. Parenthetically, it would appear that such a topic is of minor interest to non-directive counselors since they avoid making judgments in counseling relationships. Instead of avoidance, we feel that it is more effective to refine and improve the validity of judg-

---

<sup>8</sup> EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, VI (1946), 141.

ment-making in human relationships, particularly in the learning or pedagogical relationships between counselor (teacher) and student.

### *Counseling as Listening*

The fourth general characteristic of the Minnesota program has been insistence that counseling is a *mutual* process between counselor and counselee.<sup>9</sup> Minnesota counselors are not required to take the "vows of silence." Rather do they conceive their functions as somewhat similar but not identical to that of the classroom teacher who participates in the learning process, not by dominating either teaching or learning, but rather by contributing significant items of information and suggestions which are then reviewed and evaluated by the student-counselee and accepted or rejected by that same counselee. Our conception of counseling as a learning situation and experience has been that the counselor contributes information and helps the individual develop and become skilled in using methods of thinking or in bringing valid information to bear upon the solution of his own personal problems, as well as to accept emotionally himself and his situation. Darley's very significant description of his conception of counseling, *Testing and Counseling in High Schools*,<sup>10</sup> outlines at considerable length how various types of information are interpreted jointly and co-operatively by the counselor and the student so that the student may learn to appraise and to evaluate himself in terms of the demands made upon his aptitudes, interests, and personality by occupations, by the schools, and by other types of definitive situations. See also Darley's *The Interview in Counseling* and Wrenn's chapter in the *37th Yearbook*.

In contrast with this point of view we may ask certain questions of the non-directivists. Is the truth about the nature of counseling encompassed in a series of opposites, or two-order values, with no tenable ground between them? e.g., "non-direc-

<sup>9</sup> Wrenn, C. Gilbert, Ch. IV, "Counseling with Students," *37th Yearbook*, Part I, entitled *Guidance in Educational Institutions*, National Society for the Study of Education.

Darley, J. G. *The Interview in Counseling*. Washington: U. S. Dept. of Labor, 1946.

<sup>10</sup> Chicago: Science Research Associates, 1942.

tive versus directive"; "client-centered versus counselor-centered"; "silence versus advice"; "all or none," etc., etc. Or is it possible that we are dealing with continuous situations which may be appropriate or inappropriate, relevant or irrelevant, for different conditions, times, and persons?

To go on to another aspect, Rogers and his associates make much of the cardinal point, indeed it is their most important principle, that the counselee must make his own evaluations, decisions, choices, and adjustments. Such a point of view is congenial and not at all startlingly new to a Minnesota man. I know of no counselor at Minnesota who has imposed a vocational choice or any other kind of choice upon a student. Indeed, I don't see how a counselor could compel a student to choose anything. Though we avoid compulsion, we at Minnesota don't hesitate to suggest, inform, contribute, participate, help, and even advise (!) students in their attempts to understand themselves and the demands made upon them by the world about them.

At this point one may legitimately ask certain corollary questions:—Is advice identical with imposition? Can the counselor have his own judgment, express it, and yet be neutral? Is counseling restricted to therapy, to cases of emotion and confused orientation? Or is counseling also applicable to normal learning of adjustments such as choosing an occupation?

When pressed in discussion of this point, some non-directivists delimit the area of their generalizations with respect to the use of non-participating silence as a technique in counseling. Informational situations are ruled out of the scope and other restrictions made explicit. After listening to such discussions, one arrives at the supposition that what may be needed in the field of counseling is more application of the scientific method to the defining of the scope of a particular method of counseling. Is the non-directive method a generic method as well as a special technique which may be applicable to certain types of adjustments and to certain types of clients? Perhaps the answer is not "either—or" since "permissiveness" may be a universal and necessary condition of all counseling.

At the same time some of the presently defended techniques, such as "repeat-the-last-sentence," may prove to be a ritual observance rather than a necessary and universal condition for effective therapy. The careful testing of techniques to determine which are ritualistic and which are appropriate and relevant is an undertaking much to be desired by all. The unsolved major problems of counseling are, in my judgment:—"under what conditions and with what types of problems and individuals will a particular counseling technique facilitate the desired adjustment?"

To me the counselor is to be thought of as a special kind of "teacher," with primary concern for the individual's personality development, normal and deviate as well. The counselor facilitates learning, but he does not "impose" an answer to the problems faced by the student. Rather does the counselor provide knowledge of methods of problem solving and supplementary information and encouragement as well as many other facilitating contributions.

One other point concerning counseling as listening, can be dealt with only briefly in this paper. The non-directivists emphasize not only non-participation by the counselor, but also stress what I refer to as the "unfoldment" theory of how human beings learn to adjust. It is as though, as far as I can understand, all of that which is needed for growth were contained within the organism and nothing substantive need be contributed from environmental sources. If the counselor gets out of the way of the client, he can, by using certain specified techniques, make it possible for the individual to "unfold" or release that which is blocked within him. Perhaps I misunderstand, but if not, then it seems to me that we are going over the same ground that the progressive educators covered within the past decade or two. Shall the teacher get out of the classroom for fear of influencing the pupils? Shall the counselor remain silent for fear of influencing the client? I have great respect for the growth potentialities of individuals and of course development is something which only the individual can achieve for himself. Nonetheless, the individual needs assistance.

As a protest movement against the admittedly rigid impositions of some former teaching techniques, Progressive Education had its great and important influence. But some of its adherents overstated the case for "unfoldment" of the pupils' abilities through self-motivation and self-direction. We now see that the teacher can help the learning process, not merely by removing obstacles and blockings, but also equally by contributing suggestions, facts, methods, etc. Our current system of pedagogy is based upon the assumption that the teacher significantly contributes and participates in the learning process but at no time should he impose or dominate that learning process.

Now, are we again faced with a comparable "all or none" concept of participation in counseling? May not the counselor contribute anything but silence or repetition of sentences, for fear of "imposing" upon the client? If the answer is yes, then one wonders how long will be required to "go through" this all-or-none movement to a reasonable "more-or-less" position in which degree and type of participation by the counselor is determined by the type of client, problem, situation, etc., and it is no longer assumed that the counselor has no significant contribution to make other than that of establishing a catalytic condition.

#### *Other Features of Minnesota Programs*

To return from my detour through Chicago-land, a fifth general characteristic of the Minnesota program is not unique,<sup>11</sup> but at least it has been systematically explored in a preliminary manner. I refer to our early attempts to evaluate counseling, not merely by an analysis of the processes involved in counseling or by a vague assumption of outcome, but rather through experimental verification. Of the general methods used in evaluating counseling or any other personal service, the fourth is the most significant. The four general types are:

1. The hortatory appeal based upon the "reasonableness" of the method used in counseling.

<sup>11</sup> See the writings of Porter, Lewis, Snyder, and others regarding significant contributions in evaluating the non-directive methods. It should be noted that definitive studies of long-term effects, as opposed to immediate effects observable in interviews, have not been published by the non-directivists.

2. The sample and illustrative case history.
3. A critical and sometimes statistical analysis of processes of counseling.
4. An experimental determination of outcomes as measured or judged using definite criteria and using approved and conventional experimental designs modified in terms of the uniqueness of the data.<sup>12</sup>

A sixth characteristic has to do with a continuous emphasis on research, both on instruments, such as tests, high-school grades, and on other techniques and methods of analysis of aptitudes, interests, achievement, and personality. Extensive publications in the field indicate the range of the contributions of this type.

A seventh characteristic is the concept of a *balanced* student service, including the following special provisions:

1. A clinic, called the counseling bureau, in which there is a balance of different points of view and approaches, biases, and hunches. In our Counseling Bureau there are measurement experts, reading specialists, specialists in occupational adjustment and occupational information, special women counselors who are deliberately placed there to see that the woman's point of view and the woman's special interests and needs are not lost sight of even in the counseling of men, and finally "emotional counselors," including one trained in the so-called client-centered school of thought.

2. A student personnel program with specialized services including specially-trained counselors in dormitories and fraternities, specialists in speech and hearing disabilities, financial counseling, the use of activities for personal adjustment, veterans' counseling, and special counseling services for foreign students—to name a few of the many services. To single out any one part of this broad-gauge program and ignore the others, is to miss the essential emphasis upon coordination of our many types of professionalized counseling. No one type is dominant

<sup>12</sup> Williamson, E. G. and Darley, J. G. *Student Personnel Work*. New York: McGraw-Hill Book Co., 1937. Ch. IX.

Williamson, E. G. and Bordin, E. "Objective Evaluation of Student Personnel Services at the University of Minnesota." Ch. XX in *Student Personnel Services in Colleges and Universities*. Edited by John Dale Russell. Chicago: University of Chicago Press, 1941.

but rather each is an appropriate adaptation made to fit the peculiar needs of a special situation. For example, our psychologically-trained counselors employed in the dormitories certainly do not and should not counsel in the identical manner as do psychologists employed in the Student Counseling Bureau. Yet the dormitory counselors perform as important and as valid services as do those in our clinic, and in certain cases the former are more effective since they deal with students in a "real" life situation rather than in the sometimes artificial clinic interview situation.

3. An extensive program of advising and counseling in the colleges of the University as opposed to the wide technical services such as Student Counseling Bureau. Some of the faculty counselors and advisers are professionally trained in Psychology, Mental Hygiene, and related technical disciplines; whereas others are teachers of English, Mathematics, and Foreign Languages.

4. A coordinated and integrated program of consultations in marshalling the institution's entire resources to assist the individual student. No particular point of view restricts an individual counselor from using any legitimate method of counseling in dealing with a student. In this sense, Minnesota is eclectic, but I hope it is critically so of its own biases as well as of others.

### *Summary and Emphasis*

By way of summary and emphasis, counseling, *even as therapy*, seems to me to be most effective for most students when it is not restricted to that which takes place in an interview or interviews. As has been amply demonstrated in industrial personnel work, counseling is most effective when it is an integral part of a total environmental and institutional personnel program, consisting of many types of services brought to focus on the individual students' learning-needs to aid him in finding and perfecting methods of working out his own solutions to his own problems. The structuring, organizing, and coordinating of this total-environmental-learning situation is, of course, of major proportions as a task for any counselor and

personnel administrator. Continuous coordination of a multiplicity of programs and services thus becomes a dominant feature for every counselor according to the degree of his insight and skill. The emphasis is not on counseling restricted to an isolated technical clinic, but rather does a clinic become one type of counseling. Therapy in the clinical sense takes its place in a wide range of therapies provided in a student-faculty-community society. The student learns and re-learns his adjustments in this total environment and the college once more assumes its earlier societal role of facilitating the learning or development of personality.

•



# SOCIAL ADJUSTMENT AND INTEREST SCORES OF INTROVERTS AND EXTROVERTS

M. CATHARINE EVANS  
Indiana University

THIS study was attempted in order to describe the characteristics of three types of Extroverts and Introverts as measured by the *Minnesota T-S-E Inventory* in terms of performance on related sections of the *Minnesota Personality Test* and the *Kuder Preference Record*.

During the academic years, 1942 to 1945, 190 women students at Indiana University took a test battery which included the *Minnesota T-S-E Inventory*, the *Minnesota Personality Test*, and the *Kuder Preference Record*. From this group, those students who ranked in the upper and lower 25 per cent according to the norms for each of the three scores on the *Minnesota T-S-E Inventory* were chosen for this study. During the remainder of this article those students ranking in the upper 25 per cent on each of the I-E tests will be designated as Extroverts and those students ranking in the lower 25 per cent will be designated as Introverts.

## *Description of Tests*

The *Minnesota T-S-E Inventory* was constructed to measure three types of Introversion-Extroversion: Thinking, Social, and Emotional. The construction and choice of items for the three tests in the inventory were guided by the following definitions which contrast the extremes for each type of I-E:

The thinking introvert likes reflective thought, particularly that of a more abstract nature. His rational processes tend to be less dominated by objective conditions and generally accepted ideas than those of the extrovert. The thinking extrovert, however, shows a liking for overt action, and his ideas tend to be the ideas of overt actions.

The social introvert withdraws from social contacts and responsibilities and displays little interest in people. In contrast, the social extrovert seeks social contacts and depends upon them for his satisfaction.

The emotional introvert tends to repress and to inhibit the outward expression of his emotions and feelings. On the other hand, the emotional extrovert readily expresses his emotions and feelings outwardly. He shows a greater tendency to make the expected response to simple, direct emotional appeals than does the introvert.

*The Minnesota Personality Inventory* provides five separate measures of individual adjustment: Morale, Social Adjustment, Family Relations, Emotionality, and Economic Conservatism. The first four scores were included in this study as typical of measures yielded by personality tests. The authors describe the significance of these scores thus:

I. Morale. High scores are indicative of belief in society's institutions and future possibilities. Low scores usually indicate cynicism or lack of hope in the future.

II. Social Adjustment: High scores tend to be characteristic of the gregarious, socially mature individual in relations with other people. Low scores are characteristic of the socially inept or undersocialized individual.

III. Family Relations: High scores usually indicate friendly and healthy parent-child relations. Low scores suggest conflicts or maladjustments in parent-child relations.

IV. Emotionality: High scores are representative of emotionally stable and self-possessed individuals. Low scores may result from anxiety states or over-reactive tendencies.

The *Kuder Preference Record* is used extensively in the measurement of student interests. It yields scores for interest in the following nine areas: mechanical, computational, scientific, persuasive, artistic, literary, musical, social service, and clerical.

### *Method*

The four scores on the *Minnesota Personality Test* were analyzed for the Social and Emotional Introverts and Extroverts. The nine scores on the *Kuder Preference Record* were examined for the Thinking Introverts and Extroverts because

the basic definition of Thinking I-E seemed to indicate a closer relationship to measured interests than to personality characteristics. The Persuasive and Social Service scores of the *Kuder Preference Record* also were considered for the Social Introverts and Extroverts since a relationship seemed probable. Analyses of other scores on these two tests might yield interesting differences for each type of I-E, but these comparisons are left for another study.

The percentile ranks corresponding to the selected test scores were tabulated for both the Extroverts and Introverts, and the median,  $Q_1$ ,  $Q_3$ , and  $Q$  were computed. The mean raw scores also were calculated and translated to percentile ranks. These statistics were expressed in terms of percentile ranks for ease of interpretation of the differences between the two extreme groups. The critical ratios of the difference in mean raw scores to the standard error of that difference also were computed for each comparison in order to test statistically the degree of significance of the observed differences. In Table 4, the difference in mean scores and the corresponding critical ratio are presented for each difference which proved to be statistically significant, i.e., it satisfied at least the 5 per cent level of significance. In the other three tables, the statistically significant comparisons are starred.

### *Results*

#### I. For Social Extroverts and Introverts

1. They were clearly differentiated by the Social Adjustment Section of the *Minnesota Personality Test* (Table 1). The Social Introverts tended to have significantly low scores which indicate the "socially inept or undersocialized individual." Both the median and mean for the Social Introverts lay below the 20th percentile rank, and only 9, or 17 per cent, of their scores were above the 46th percentile rank for the norm group. On this same section the Social Extroverts tended to have high scores which are characteristic of the "gregarious, socially mature individual." Both the median and mean for

The social introvert withdraws from social contacts and responsibilities and displays little interest in people. In contrast, the social extrovert seeks social contacts and depends upon them for his satisfaction.

The emotional introvert tends to repress and to inhibit the outward expression of his emotions and feelings. On the other hand, the emotional extrovert readily expresses his emotions and feelings outwardly. He shows a greater tendency to make the expected response to simple, direct emotional appeals than does the introvert.

*The Minnesota Personality Inventory* provides five separate measures of individual adjustment: Morale, Social Adjustment, Family Relations, Emotionality, and Economic Conservatism. The first four scores were included in this study as typical of measures yielded by personality tests. The authors describe the significance of these scores thus:

I. Morale: High scores are indicative of belief in society's institutions and future possibilities. Low scores usually indicate cynicism or lack of hope in the future.

II. Social Adjustment: High scores tend to be characteristic of the gregarious, socially mature individual in relations with other people. Low scores are characteristic of the socially inept or undersocialized individual.

III. Family Relations: High scores usually indicate friendly and healthy parent-child relations. Low scores suggest conflicts or maladjustments in parent-child relations.

IV. Emotionality: High scores are representative of emotionally stable and self-possessed individuals. Low scores may result from anxiety states or over-reactive tendencies.

The *Kuder Preference Record* is used extensively in the measurement of student interests. It yields scores for interest in the following nine areas: mechanical, computational, scientific, persuasive, artistic, literary, musical, social service, and clerical.

### *Method*

The four scores on the *Minnesota Personality Test* were analyzed for the Social and Emotional Introverts and Extroverts. The nine scores on the *Kuder Preference Record* were examined for the Thinking Introverts and Extroverts because

the basic definition of Thinking I-E seemed to indicate a closer relationship to measured interests than to personality characteristics. The Persuasive and Social Service scores of the *Kuder Preference Record* also were considered for the Social Introverts and Extroverts since a relationship seemed probable. Analyses of other scores on these two tests might yield interesting differences for each type of I-E, but these comparisons are left for another study.

The percentile ranks corresponding to the selected test scores were tabulated for both the Extroverts and Introverts, and the median,  $Q_1$ ,  $Q_3$ , and  $Q$  were computed. The mean raw scores also were calculated and translated to percentile ranks. These statistics were expressed in terms of percentile ranks for ease of interpretation of the differences between the two extreme groups. The critical ratios of the difference in mean raw scores to the standard error of that difference also were computed for each comparison in order to test statistically the degree of significance of the observed differences. In Table 4, the difference in mean scores and the corresponding critical ratio are presented for each difference which proved to be statistically significant, i.e., it satisfied at least the 5 per cent level of significance. In the other three tables, the statistically significant comparisons are starred.

### Results

#### I. For Social Extroverts and Introverts

1. They were clearly differentiated by the Social Adjustment Section of the *Minnesota Personality Test* (Table 1). The Social Introverts tended to have significantly low scores which indicate the "socially inept or undersocialized individual." Both the median and mean for the Social Introverts lay below the 20th percentile rank, and only 9, or 17 per cent, of their scores were above the 46th percentile rank for the norm group. On this same section the Social Extroverts tended to have high scores which are characteristic of the "gregarious, socially mature individual." Both the median and mean for

the Extroverts reached the 70th percentile rank and only 6, or 16 per cent, of them were below the norm median. The critical ratio of the difference in mean raw scores was 8.11, which indicated less than one

TABLE 1

*Analysis of Percentile Ranks on 4 Sections of the Minnesota Personality Test and 2 Sections of the Kuder Preference Record for 37 Social Extroverts and 52 Social Introverts*

	Minnesota Personality Test					
	I Morale		II Social*		III Family	
	Extroverts	Introverts	Extroverts	Introverts	Extroverts	Introverts
Statistic	<i>N</i> = 37	<i>N</i> = 52	<i>N</i> = 37	<i>N</i> = 52	<i>N</i> = 37	<i>N</i> = 52
$Q_n$	88	75	90	38.5	80	73
Median	51	41	70	17	54	55.5
$Q_1$	35	20	55	6	37	18
$Q$	26.5	27.5	17.5	16.4	21.5	27.5
Mean	62	46	72	19	54	45

	Minnesota Personality Test		Kuder Preference Record			
	IV Emotionality*		4 Persuasive*		8 Social Service	
	Extroverts	Introverts	Extroverts	Introverts	Extroverts	Introverts
Statistic	<i>N</i> = 37	<i>N</i> = 52	<i>N</i> = 35	<i>N</i> = 51	<i>N</i> = 35	<i>N</i> = 51
$Q_n$	70	69	85	51	89	90
Median	47	30.5	62	31	75	74
$Q_1$	31	8	33	2	42	32
$Q$	19.5	30.5	26	24.5	23.5	29
Mean	50	31	63	27	67	67

chance in 100 that the obtained difference could be due to chance errors of sampling (Table 4).

- On the Emotionality Section of the *Minnesota Personality Test*, the Extroverts resembled the original norm group, but the mean, median, and first quartile points for the Introverts were low. The Social Introverts seemed to display "anxiety states and over-reactive tendencies." The difference in mean scores for the Introvert and Extrovert

was statistically significant at the 5 per cent level (Table 4).

3. The differences between the Social Introverts and Extroverts on the other two sections of the *Minnesota Personality Test* were not statistically significant. Both groups had medium scores on the Morale and Family Relation Sections.
4. The Social Extroverts and Introverts were differentiated by the Persuasive Score on the *Kuder Preference Record*, but not by the Social Service Score. The Introverts ranked low in Persuasive interest with a mean score corresponding to the 27th percentile rank. The Extroverts in contrast displayed a greater degree of interest in Persuasive activities with a mean score corresponding to the 63rd percentile rank. A critical ratio of 4.06 indicates that the difference in means for the Extroverts and Introverts was statistically significant (Table 4).

A similar contrast was expected for the Social Service Scale, but both the Extroverts and Introverts ranked above average on it with equal median and equal mean scores. Kuder describes the Persuasive Scale as including "activities in the general field of convincing, persuading, managing, and selling" and the Social Service Scale as "activities indicative of an interest in people with the emphasis on those which involve personally helping people in a fairly direct way." Kuder reports such low intercorrelation coefficients for those scales that it must be concluded that they measure very different interests. Apparently, both the Social Extroverts and Introverts rank above average in interest in Social Service activities, but only the Social Extroverts are interested in Persuasive activities.

## II. For Emotional Extroverts and Introverts

1. The definitions of Emotional I-E for the *Minnesota T-S-E Inventory* and Emotionality for the *Minne-*

*sota Personality Test* differ so greatly that it was difficult to predict the results on the Emotionality Section for the Emotional Extroverts and Introverts. The Introverts resembled the original norm group with both the median and mean scores corresponding to the 50th percentile rank (Table 2). The Extroverts had a significantly lower mean Emotionality score than the Introverts (Table 4). Both measures of central tendency for the Extroverts were below the 39th percentile rank, and the middle

TABLE 2

*Analysis of Percentile Ranks on 4 Sections of the Minnesota Personality Test for 38 Emotional Extroverts and 50 Emotional Introverts*

Statistic	Minnesota Personality Test							
	I Morale		II Social		III Family		IV Emotionality*	
	Extroverts	Introverts	Extroverts	Introverts	Extroverts	Introverts	Extroverts	Introverts
N	N=38	N=50	N=38	N=50	N=38	N=50	N=38	N=50
$Q_n$	77.5	80	75	77.5	80.5	90	58.5	77.5
Median	52.5	50.5	55	48.5	59.5	65	38.5	49.5
$Q_1$	28.5	23.5	22	26.5	33	29.5	11	25.5
$Q$	24.5	28+	26.5	25.5	23.7	30+	23.7	26
Mean	54	49	46	48	49	56	34	50

50 per cent of their scores lay between the 11th and 58th percentile ranks. According to the *Minnesota Personality Inventory*, the Emotional Extroverts were less emotionally stable than Introverts and they showed a greater tendency to have "anxiety states and over-reactive tendencies."

- The differences for the other three sections of the *Minnesota Personality Test* were not significant. Both the Emotional Extroverts and Introverts had medium measures of central tendency for the Morale, Social, and Family Relation Sections.

### III. For Thinking Extroverts and Introverts

From the basic definition of Thinking I-E, differences in the measured interests of introverts and extroverts



were expected. The results for each section of the *Kuder Preference Record* are as follows:

### 1. Literary Activities

There was a striking contrast in the interest of Introverts and Extroverts in these activities. The Extroverts ranked significantly lower in Literary interest than the Introverts. Both the median and

TABLE 3

*Analysis of Percentile Ranks on 9 Sections of the Kuder Preference Record for 38 Thinking Introverts and 38 Thinking Extroverts*

	<i>Kuder Preference Record</i>					
	I Mechanical		II Computational		III Scientific	
	Extroverts	Introverts	Extroverts	Introverts	Extroverts	Introverts
$Q_3$	78	75	57	62.5	61	79.5
Median	57	39.5	43	49	40	36.5
$Q_1$	35.5	21	30	27	14	14
$Q$	21+	27	13.5	17.7	23.5	32.7
Mean	60	50	43	49	38	52

	<i>Kuder Preference Record</i>					
	IV Persuasive*		V Artistic		VI Literary*	
	Extroverts	Introverts	Extroverts	Introverts	Extroverts	Introverts
$Q_3$	68	51	80	89	35	93
Median	48	35	67	76	14	55.5
$Q_1$	29	7	34	24	5	33
$Q$	19.5	22	23	32.5	15	30
Mean	49	30	68	72	15	68

	<i>Kuder Preference Record</i>					
	VII Musical		VIII Social Service*		IX Clerical	
	Extroverts	Introverts	Extroverts	Introverts	Extroverts	Introverts
$Q_3$	69	72	92	81	70	56.5
Median	32	37	81	54	36	34
$Q_1$	11	18	53	22.5	23	9
$Q$	29	27	19.5	29+	23.5	23.7
Mean	34	41	81	55	46	35

mean scores of the Extroverts lay in the lower 15 per cent of the norm group (Table 3). In fact, all but two of the Extrovert scores fell below the norm median. In contrast, the measures of central tendency for the Introverts were above the median, and 25 per cent of them had scores in the upper 7 per cent of the norm group. The Extroverts were less interested in Literary activities than the Introverts,

TABLE 4

*Differences in Mean Raw Scores and Critical Ratios for Statistically Significant Comparisons in This Study*

	Difference <sup>†</sup> in mean raw scores	Critical ratio	Level of Signifi- cance
<i>Minnesota Personality Test</i>			
II Social			
Social Extroverts & Introverts ....	+38.94	8.11	1%
IV Emotionality			
Social Extroverts & Introverts .....	+12.28	2.61	5%
Emotional Extroverts & Introverts .	-10.83	2.21	5%
<i>Kuder Preference Record</i>			
Persuasive			
Social Extroverts & Introverts .....	+13.81	4.06	1%
Thinking Extroverts & Introverts .....	+6.90	2.16	5%
Literary			
Thinking Extroverts & Introverts .....	-23.01	6.06	1%
Social Service			
Thinking Extroverts & Introverts .....	+12.52	3.34	1%

†+ indicates a higher mean score for the Extroverts and - indicates a higher mean score for the Introverts.

and this difference seemed in accord with the original definition of Thinking I-E.

## 2. Social Service Activities

The Thinking Extroverts ranked high in interest in these activities. The median and mean scores for the Extroverts corresponded to the 81st percentile rank, and 75 per cent of them had scores above the norm median. The distribution of scores of the Introverts resembled closely the norm group and they showed only a medium degree of Social Service interest. The difference between the mean

scores of the Extroverts and Introverts was statistically significant. This interest displayed by the Extroverts in contrast to the Introverts seemed in accord with the original definition of Thinking I-E.

3. Persuasive Activities

The Thinking Introverts ranked low in Persuasive interest. The mean score for the Introverts equaled the 30th percentile rank, and approximately 75 per cent of their scores were below the norm median (Table 3). The Extroverts showed a medium degree of interest in Persuasive activities, and they had a significantly higher mean score than the Introverts. The Introvert lack of interest in these activities seemed in accord with the original definition of Thinking I-E.

4. Clerical Activities

The Introverts had a lower mean Clerical score than the Extroverts, but the difference was not statistically significant. The median and mean scores of the Introverts corresponded to the 34th and 35th percentile ranks, respectively, and 75 per cent of their scores were in the lower 57 per cent of the norm group. A lack of interest in Clerical Activities for the Introverts was expected from the definition of Thinking I-E, but in this study a significant difference between the Introverts and Extroverts was not found.

5. Mechanical Activities

The Extroverts showed a greater interest in mechanical activities than the Introverts, but the difference was not significant. The median for the Introverts was below the norm median, but their mean score corresponded exactly to this median. Both the median and mean scores for the Extroverts were above the norm median. A significantly greater mechanical interest for the Extroverts in contrast to the Introverts was expected from the original definition of Thinking I-E.

#### 6. Artistic Activities

Previous research results with the *Minnesota T-S-E Inventory* indicated that the Thinking Introverts would rank higher in Artistic interest than the Extroverts. In this study both groups tended to have high scores although the measures of central tendency for the Introverts were higher than those of the Extroverts. The difference between the Introverts and Extroverts was not significant, but the direction of the difference was in accord with expectation.

#### 7. Scientific Activities

It was difficult to predict results from the original definition of Thinking I-E, and no significant difference was found between the Introverts and Extroverts in Scientific interest. The medians were approximately equal but the mean score of the Extroverts was lower than that of the Introverts. Male Extroverts might show significantly less interest in scientific activities than male Introverts.

#### 8. Musical Activities

Differences similar to those found for the Literary Activities were expected. Actually both the Introverts and Extroverts ranked below the median in musical interest although the mean and median of the Introverts were slightly higher than the corresponding scores for the Extroverts.

#### 9. Computational Activities

As with the Scientific Activities it was difficult to predict results on the Computational Scale. There was not a significant difference between the two extreme groups, but both measures of central tendency for the Introverts were higher than the corresponding one for the Extroverts.

### *Summary*

1. A close relationship was found between Social I-E and Social Adjustment as measured by the *Minnesota Personality*

*Test.* Social Extroverts had high scores which are characteristic of the "gregarious, socially mature individual," and the Introverts had low scores which indicate undersocialization.

2. Both the Social and Emotional Introverts and Extroverts were differentiated by the Emotionality Section of the *Minnesota Personality Test*. The Social Introverts and Emotional Extroverts tended to have low scores which indicate "anxiety states and over-reactive tendencies." In contrast, the significantly higher mean scores of the Social Extroverts and Emotional Introverts signified satisfactory emotional adjustment.

3. There was a significant difference in the interest in Persuasive activities for the Social Extroverts and Introverts. The Introverts ranked low in Persuasive interest in contrast to a higher degree of interest displayed by the Extroverts. Both the Social Extroverts and Introverts showed a high degree of interest in Social Service Activities.

4. The results with the *Kuder Preference Record* revealed contrasts in the interests of the Thinking Extroverts and Introverts. Significant differences were found for the Literary, Social Service, and Persuasive scores. The Extroverts ranked low in interest in Literary activities and high in Social Service interest in contrast to a medium degree of interest in both types of activities for the Introverts. The Introverts ranked low in interest in Persuasive activities while the Extroverts showed a medium degree of Persuasive interest. The other differences were not statistically significant, but there was some evidence that the Extroverts ranked higher in Mechanical and Clerical interests than the Introverts while the Introverts were more interested in Scientific Artistic, Musical, and Computational activities than the Extroverts.



## MEMBERS OF THE AMERICAN COLLEGE PERSONNEL ASSOCIATION

- Adams, Dr. Clifford R., Marriage Counseling Service, Pennsylvania State College, State College, Pennsylvania
- Adams, Dr. William M., c/o Evelyn Brandt, 5503 Fifteenth Avenue, N.E., Seattle, Washington
- Aiken, Prof. N. J., Director of Placement Bureau Association, Professor of Economics, State College of Washington, Pullman, Washington
- Albright, Miss Thelma, Dean of Students, Queens College, Charlotte, North Carolina
- Alexander, Miss Ruth Lenore, 804½ E. 8th Street, Bloomington, Indiana
- Allen, Cecil H., Head of the Division of Personnel Service, N.E. Missouri State Teachers College, Kirksville, Missouri
- Anderson, Gordon V., 201 Fayweather Annex, Northwestern University, Evanston, Illinois
- Andrews, Dr. Elizabeth G., Director of Personnel, Florida State College for Women, Tallahassee, Florida
- Applegate, C. William, 7740 N. Ashland, Chicago 26, Illinois
- Austin, Sydney F., 34 Crestview Road, Mountain Lakes, New Jersey
- Babcock, Miss Fern, National Student Secretary, Y.W.C.A., 600 Lexington Avenue, New York, N. Y.
- Bagley, Miss Leila, 1000 Morningside Drive, New York 27, N. Y.
- Bailey, Prof. Harold W., Director of Personnel Bureau, University of Illinois, Urbana, Illinois
- Bathurst, Dr. J. E., National Association of Foremen, 11 W. Monument Bldg., Dayton Ohio
- Baum, Dean Paul B., Colorado Women's College, 18th and Poplar Street, Denver 7, Colorado
- Baxt, Roland, Executive Director, Federation Employment Service, 67 W. 47th St., New York, N. Y.
- Belknap, Miss A. Fredericka, Director of Personnel Bureau, New Jersey College for Women, New Brunswick, New Jersey
- Beery, George S., Dean of Students, Drake University, Des Moines 11, Iowa
- Bell, Dr. Hugh M., Student Personnel and Guidance, Chico State College, Chico, California
- Bent, Leo G., Assistant Professor of Education, Western Illinois State Teachers College, Macomb, Illinois
- Berg, Dr. Irwin A., Clinical Counselor and Associate in Psychology, University of Illinois, Urbana, Illinois

- Bergstresser, Dr. John L., 202 Reynolds Club, University of Chicago, Chicago 37, Illinois
- Bernard, Lloyd D., Mgr., Bureau of Guidance and Placement, University of California, 102 Administration Bldg., Berkeley 4, Calif.
- Bigelow, Miss Mary D., Chairman of Advising, Stephens College, Columbia, Missouri
- Billings, Earle Montoe, Personnel Director, Eastman Kodak Company, 343 State Street, Rochester, New York
- Bishop, Mrs. Joan Fiss, Director of Placement Office, Wellesley College, Wellesley 81, Massachusetts
- Blaesser, Dr. Willard W., Director of Student Personnel Services, Montana State University, Missoula, Montana
- Blanchard, Dr. (Miss) Leslie, The Packer Collegiate Institute, 170 Joralemon Street, Brooklyn 2, N. Y.
- Blanding, Sarah G., President, Vassar College, Poughkeepsie, New York
- Bonney, Dr. M. E., Guidance and Personnel, North Texas State Teachers College, Denton, Texas
- Borow, Dr. Henry, The General College, University of Minnesota, Minneapolis 14, Minnesota
- Boyer, Carl Wright, Professor of Education and Psychology, Muhlenberg College, Allentown, Pa.
- Boyles, Lt. R. E., 532 River Street, Hoboken, New Jersey
- Brayfield, Dr. Arthur Hills, Dean of Student Personnel, Long Beach City College, Long Beach, California
- Brett, Miss Ruth, Acting Personnel Director, P. O. Box 445, Tuskegee Institute, Tuskegee, Alabama
- Bridgman, Donald S., American Telephone and Telegraph Company, 195 Broadway, New York, N. Y.
- Bridgman, Pres. Ralph B., Hampton Institute, Hampton, Virginia
- Brockman, Dr. L. O., Director, Testing and Counseling, Montana State College, Bozeman, Montana
- Brooks, Mrs. Dorothy, Dean of Women, Dennison University, Granville, Ohio
- Brottemarkle, Prof. Robert A., Personnel Officer, 116 College Hall, University of Pennsylvania, Philadelphia, Pennsylvania
- Brouwer, Paul J., 405 S. 8th Avenue, La Grange, Illinois
- Brown, Charles K., Director of Student Affairs and Publications Editor, Alabama Polytechnic Institute, Auburn, Alabama
- Brown, Kirtley, Director of Personnel, Alabama Polytechnic Institute, Auburn, Alabama
- Brown, Mrs. Lucille B., Chairman, Teacher Education Department, Child Education Foundation, 535 E. 84th St., New York, N. Y.
- Brumbaugh, Dr. A. J., American Council on Education, 744 Jackson Place, Washington 6, D. C.
- Brummett, Dr. Carl A., Director of Student Activities, Jefferson College, 1528 Locust Street, St. Louis, Missouri
- Buller, Dr. Francis Paul, Director of Testing and Counseling, Evansville College, Evansville, Indiana



- Bunn, John W., Springfield College, Springfield, Massachusetts  
Burdin, L. Gray, Chairman of Men's Council, Butler University, Indianapolis 7, Indiana  
Burnham, Dr. Paul S., Associate Director of Student Appointment Bureau, Yale University, New Haven, Connecticut  
Butler, Miss Katherine M., Assistant Director of Placement and Service Bureau, Louisiana Polytechnic Institute, Ruston, Louisiana  
Butts, Miss Ruth E., Dean of Women, Hillsdale College, Hillsdale, Michigan  
Callis, Robert, Research Assistant in Clinical Counseling, General College, University of Minnesota, Minneapolis, Minn.  
Cameron, Mrs. Marian Lippitt, University Testing Service, Syracuse University, Syracuse, N. Y.  
Camp, Miss Frances M., Director of Bureau of Appointments, State University of Iowa, Iowa City, Iowa  
Cannon, Dean C. W., Consulting Psychologist, 4567 Grand Ave., Western Springs, Illinois  
Carpenter, Miss M. Helen, Acting Director of Placement Bureau, University of Colorado, Boulder, Colorado  
Cashman, Robert D., Assistant Director Placement, Rhode Island State College, Kingston, Rhode Island  
Celine, Sister Mary, Immaculata College, Immaculata, Pennsylvania  
Chalfen, Leo, 7342 67 Drive, Middle Village, Long Island, N. Y.  
Chambers, Dr. O. R., Head of Department of Psychology, Oregon State College, Corvallis, Oregon  
Clark, Miss Barbara M., Assistant Director of Student Activities Bureau, University of Minnesota, Minneapolis, Minnesota  
Clevenger, John C., Dean of Students, Colorado A. and M. College, Fort Collins, Colorado  
Coates, Dorothy T., Director, Katherine Gibbs School, 720 N. Michigan Ave., Chicago, Ill.  
Colvin, Harold William, Y.M.C.A. National Council, 19 S. LaSalle Street, Chicago, Illinois  
Compton, Dean Richard K., 521 W. Green Street, Hastings, Michigan  
Congdon, Dr. Nora A., Personnel Research Officer, Colorado State College of Education, Greeley, Colorado  
Congdon, Dean Wray H., Dean of Undergraduates, Lehigh University, Bethlehem, Pennsylvania  
Conwell, Dean H. H., Beloit College, Beloit, Wisconsin  
Corson, Dean James Hunt, Vice-Dean, Stockton Junior College, Stockton, California  
Crawley, Dr. S. L., Director of Department of Student Personnel, Colorado State College of Education, Greeley, Colorado  
Crenshaw, Mrs. James L., Director of Bureau of Recommendations, Bryn Mawr College, Bryn Mawr, Pennsylvania  
Croft, Dr. Lysle W., Director of Guidance, The University Personnel Office, University of Kentucky, Lexington, Kentucky  
Crosby, Dr. Richard C., Director of Student Counseling, Miami University, Oxford Ohio

- Culley, Benjamin Hays, Dean of Men, 1600 Campus Rd., Los Angeles 41, Calif.
- Cuninggim, Miss Margaret L., Hockaday Junior College, Dallas 6, Texas
- Dalton, Glen D., N. Y. University Christian Association, 26 Washington Place, Press Annex, Second Floor, New York 3, N. Y.
- Darley, Dr. John G., 101 Eddy Hall, University of Minnesota, Minneapolis 14, Minnesota
- Davies, Roy O., Director of Personnel, Western Auto Supply Company, 2107 Grand Avenue, Kansas City, Missouri
- Davis, Victor M., Director of Bureau of Appointments, University of Tennessee, Knoxville, Tennessee
- Davidson, Robert L. D., Placement Bureau, Broad St. and Montgomery Ave., Philadelphia 22, Pa.
- Dawson, J. D., Personnel Director, Antioch College, Yellow Springs, Ohio
- Deabler, Dr. Herdis L., Personnel Director, North Central College, Naperville, Illinois
- Degolyer, Mrs. Robert S., 125 Kedzie Street, Evanston, Illinois
- Densmore, Harvey B., Chairman of General Studies, University of Washington, Seattle, Washington
- DeVoss, Dr. James C., Dean of Upper Division, San Jose State College, San Jose, California
- Diggs, Rachel, Director, Educator, Guidance and Research, Winston-Salem Teachers College, Winston-Salem 6, North Carolina
- Dillman, Pres. Grover C., Michigan College of Mining and Technology, Houghton, Michigan
- Donahue, Dr. Wilma T., Department of Psychology, University of Michigan, Ann Arbor, Michigan
- Dooley, Raymond Nelson, Director of Student Personnel Services, Illinois Wesleyan University, Bloomington, Ill.
- Doty, Miss Katherine S., Assistant to the Dean, Barnard College, Columbia University, New York, N. Y.
- Dozier, Prof. Doris, Recorder and Secretary of Placement, Mills College, Mills College, California
- Dreffin, William B., 1714 N.E. Lincoln St., Minneapolis, Minnesota
- Dressler, Miss Marguerite R., Assistant Director of Personnel, Florida State College for Women, Tallahassee, Florida
- Drought, Dr. Neal E., RCA Victor Division, Personnel Administration, Camden, New Jersey
- Dugan, Willis E., Director of Student Personnel, University High School, University of Minnesota, Minneapolis, Minnesota
- Duggan, L. F., Registrar, Michigan College of Mining and Technology, Houghton, Michigan
- Dunham, Charles Vernon, Director of Student Employment, University of Texas, Austin 12, Texas
- Dysinger, Dr. Wendall S., Dean, MacMurray College, Jacksonville, Illinois
- Edwards, Miss Marcia, Professor of Education and Assistant to the

- Dean, College of Education, University of Minnesota, Minneapolis, Minnesota
- Eldersveld, Miss Wilma, Psychological Clinic, 1027 E. Huron Street, University of Michigan, Ann Arbor, Michigan
- Elias, Virginia, Director of Student Aid and Employment, Occidental College, 1600 Campus Rd., Los Angeles 19, Calif.
- Ellingson, Pres. Mark, Rochester Institute of Technology, 65 South Plymouth Avenue, Rochester 8, New York
- Erickson, Kenneth, Assistant in Student Affairs Office, Washington State College, Pullman, Washington
- Evans, John J., Jr., Armstrong Cork Company, Lancaster, Pennsylvania
- Evans, Miss Mary Catherine, Vocational Adviser for Women, Indiana University, Bloomington, Indiana
- Ewan, Mrs. Mary A., Associate Chairman, Appointments Division, College of Education, Arps Hall, Ohio State University, Columbus, Ohio
- Failing, Miss Jean, New York State College of Home Economics, Cornell University, Ithaca, New York
- Falvey, Frances, Adviser to Freshmen, Hollins College, Virginia
- Faries, Miss Miriam, Assistant Dean of Students, College of the City of New York, 139th and Convent Avenue, New York, N. Y.
- Farrington, E. H., Director of Placement Service, North Texas State Teachers College, Denton, Texas
- Feder, Dr. Daniel D., Dean of Students, University of Denver, Denver, Colorado
- Feldt, Miss Irene C., Director of Placement for Women, Purdue University, La Fayette, Indiana
- Felsted, Mrs. Leona Wise, Dean of Women, Ohio Wesleyan University, Delaware, Ohio
- Fisher, Prof. Charles A., Director of Education Placement Bureau, Temple University, Broad and Montgomery, Philadelphia, Pennsylvania
- Fisk, Miss Helen G., Associate Director, Western Personnel Institute, 30 North Raymond Avenue, Pasadena, California
- Flood, Julian J., Personnel Director, Tuskegee Institute, P. O. Box 445, Tuskegee, Alabama
- Focht, Miss Helen E., Old Capitol, State University of Iowa, Iowa City, Iowa
- Foley, John D., Assistant to the Dean, Office of Dean of Students, University of Minnesota, Minneapolis 14, Minnesota
- Fox, H. Clifford, University of Dubuque, Dubuque, Iowa
- Gaetje, Miss Lucille, Director of Personnel for Women, Owens-Illinois Glass Company, Gas City, Indiana
- Gallagher, Dr. Marie K., Associate Dean, Bureau of Educational Guidance, Hunter College, 695 Park Avenue, New York 21, N. Y.
- Garrett, Mrs. Mary, Director of Admissions and Student Personnel, Bennington College, Bennington, Vermont

- Garvey, Beth Porter, Dean of Women, State Teachers College, St. Cloud, Minn.
- Gebauer, Miss Dorothy, Dean of Women, University of Texas, Austin 12, Texas
- Gerber, Joe N., Director of Student Personnel, Natchitoches, Louisiana
- Gerken, Clayton d'a, Department of Psychology, State University of Iowa, Iowa City, Iowa
- Gerth, Mrs. Josephine Hammond, Hunter College, 695 Park Avenue, New York, N. Y.
- Gilbert, Dr. William M., Clinical Counselor, Personnel Bureau, University of Illinois, Urbana, Illinois
- Gillingham, J. B., Department of Economics, Wayne University, Detroit 1, Michigan
- Gire, Myron E., Director of Personnel, Wittenberg College, Springfield, Ohio
- Goetch, E. W., Director of Placement Bureau, Iowa State Teachers College, Cedar Falls, Iowa
- Gordy, Miss Margaret D., Placement Director, Wilson College, Chambersburg, Pennsylvania
- Gould, Joseph E., Acting Director of Personnel, Colorado A. and M. College, Fort Collins, Colorado
- Gregory, Dr. Wilbur S., Admissions Office, University of New Mexico, Albuquerque, New Mexico
- Griffen, Miss Alice J., Director of Personnel Service, Wright Junior College, 3400 North Austin Avenue, Chicago, Illinois
- Gum, Dr. Wanda N., 5811 West Erie, Chicago 44, Illinois
- Gus, Charles Edward, Executive Secretary, College of Engineering, New York University, University Heights, New York, N. Y.
- Haggerty, William J., President, State Teachers College, New Paltz, New York
- Hale, Dr. Lincoln B., President, Evansville College, Evansville 4, Indiana
- Hand, Thomas J., Saranac Lake Study and Craft Guild, 100 Main Street, Saranac Lake, New York
- Hanson, Miss Anna M., Simmons College, Boston, Massachusetts
- Hanson, Ernest E., Dean of Men, Northern Illinois State Teachers College, De Kalb, Illinois
- Hauptman, Leo M., Registrar and Director of Student Personnel, Kalamazoo College, Kalamazoo 49, Michigan
- Hausam, Miss Winifred M., Director of Western Personnel Service, 30 North Raymond Avenue, Pasadena, California
- Hawkes, Mrs. Herbert E., Dean of Students, Mills College, Oakland 13, California
- Held, Dr. O. C., 7 Briar Cliff Road, Ben Avon Heights, Pittsburgh 2, Pennsylvania
- Helser, Dean Maurice David, Personnel Director, Iowa State College, Ames, Iowa
- Henderson, Mrs. Virginia Kinsman, Personnel Officer, 19 West Plumstead Avenue, Lansdowne, Pennsylvania

- Henley, Dr. W. Ballentine, President, College of Osteopathic Physicians, 1721 Griffen Avenue, Los Angeles, California
- Henry, Dr. Edwin R., A.G.O., 270 Madison Ave., New York, N. Y.
- Heston, Dr. Joseph C., Director of Bureau of Testing and Research, DePauw University, Greencastle, Indiana
- Hicks, Arthur C., Western Washington College of Education, Bellingham, Washington
- Hill, George E., Head of Department of Education and Director of Teacher Training, Kansas State Teachers College, Emporia, Kansas
- Hill, Jr., Prof. J. Lawrence, Department of Mechanical Engineering, University of Rochester, Rochester, N. Y.
- Hilliard, George, Director of Personnel and Guidance, Western Michigan College of Education, Kalamazoo, Michigan
- Hilton, Miss M. Eunice, Dean of Women, Syracuse University, Syracuse, N. Y.
- Hoag, R. G., Registrar, The New York State Institute of Applied Arts and Sciences, Utica, N. Y.
- Holcomb, Miss Leila L., Assistant Counselor, University Residences for Women, University of Texas, Austin 12, Texas
- Holkje, John E., Registrar, Western Michigan College of Education, Kalamazoo, Michigan
- Holmes, John L., Assistant Director of Personnel, Iowa State College, Ames, Iowa
- Holmes, Dr. Lulu H., Dean of Women, State College of Washington, Pullman, Washington
- Homan, Walter J., Dean of Student Personnel, San Francisco State College, 124 Buchanan Street, San Francisco 2, California
- Hopkins, E. H., Assistant to the President, State College of Washington, Pullman, Washington
- Hoppock, Prof. Robert, Department of Education, New York University, New York, N. Y.
- Howard, Dean Charles W., Lewis and Clark College, 0615 S. W. Palatine Hill Road, Portland 1, Oregon
- Humphreys, Dr. J. Anthony, Bureau of Occupations, Hunter College, 695 Park Ave., New York, N. Y.
- Hupp, James L., Dean of Students, West Virginia Wesleyan College, Buckhannon, West Virginia
- Ingalls, Les W., Acting Director of Admissions and Personnel for Men, Middlebury College, Middlebury, Vermont
- Jason, William Barrington, Vice-Pres., Dean of Students, Lincoln University, Jefferson City, Missouri
- Jeep, Henry Adelbert, Assistant Professor of Education, Ball State Teachers College, Muncie, Indiana
- Johnson, Albert P., 1625 Perry Street, Columbus 1, Ohio
- Johnson, Clyde Sanfred, 202 Administration Building, U.C.L.A., 405 Hilgard Avenue, Los Angeles 24, California

- Johnson, Theron A., 209 Eddy Hall, University of Minnesota, Minneapolis 14, Minnesota
- Johnson, Prof. William C., Palmyra College, Diamond, Ohio
- Jones, Dr. Edward S., Director of Personnel Research, University of Buffalo, Buffalo, New York
- Jones, Miss Katherine J., Executive Secretary, Nursery Training School of Boston, 355 Marlborough Street, Boston 15, Massachusetts
- Jones, Dr. Lonzo, Coordinator of Student Personnel Services, Indiana State Teachers College, Terre Haute, Indiana
- Jox, Marshall J., Personnel Director, Valparaiso University, Valparaiso, Indiana
- Kennedy, Charles E., Dean of Men, West Virginia Institute of Technology, Montgomery, West Virginia
- Kennedy, Thomas M., Director of Student Personnel, Loyola University, 6525 Sheridan Road, Chicago 26, Illinois
- Kennon, Miss Ann Byrd, Personnel Director for Women, University of Colorado, Boulder, Colorado
- Kent, Dean Edwin, Dean of Men, Santa Rosa Junior College, Santa Rosa, California
- Kirkpatrick, Dean Forrest H., Bethany College, Bethany, West Virginia
- Kittrell, Miss Flemma P., Dean, Home Economics Division, Howard University, Washington 1, D. C.
- Knapp, A. Blair, Dean of Students, Temple University, Philadelphia 22, Pa.
- Kohn, Jr., Nathan, Veterans Counseling, Washington University, St. Louis 5, Missouri
- Kopas, Joseph S., Director of Department of Development, Fenn College, Cleveland, Ohio
- Kuder, G. Frederic, 306 Edmonds Bldg., 917 15th St., Washington 5, D. C.
- LeBarre, Miss Corrine G., Research Assistant, Western Personnel Service, 30 North Raymond Avenue, Pasadena, California
- Lahr, Miss Patricia A., Director of Student Union, University of Nebraska, 14th and R, Lincoln, Nebraska
- Lane, Mrs. Estella Hitchcock, Dean of Students, Frances Shimer College, Mount Carroll, Illinois
- Larsen, Robert P., Outlook Sanitorium, Urbana, Illinois
- Lauer, Dean George N., Dean of Men, Central Michigan College of Education, Mt. Pleasant, Michigan
- Leetch, Miss Florence K., Appointments Office, Harvard University, Cambridge 38, Massachusetts
- Leetch, George N. P., Director of College Placement Service, The Pennsylvania State College, State College, Pennsylvania
- Lemon, Dean Erwin B., Dean of Admissions, Oregon State College, Corvallis, Oregon
- Lemmon, William Burton, Dir., University Guidance Service, Faculty Exchange, University of Oklahoma, Norman, Oklahoma

- Leo, Brother Ignatius, Christian Brothers College, East Parkway at Central, Memphis 4, Tennessee
- Lindsey, Gardner L., Department of Social Relations, Harvard University Cambridge 38, Mass.
- Lloyd-Jones, Prof. Esther, Associate Director of Personnel, Teachers College, Columbia University, New York, N. Y.
- Loescher, Frank Samuel, Secretary, Placement Service, American Friends Service Comm., 20 South 12th St., Philadelphia 7, Pa.
- Lovell, Dr. George D., Grinnell College, Grinnell, Iowa
- Luehring, Ellen D., Head, Personnel Service Bureau, Division of Higher Education, Board of Education of Presbyterian U.S.A. Church, 804 Witherspoon Bldg., Philadelphia, Pa.
- Lyman, E., Chairman, Board of Personnel Administration, Northwestern University, 102 Lunt Building, Evanston, Illinois
- Lynn, Miss Virginia, Head, Student Personnel and Guidance Testing Service, University of Wyoming, Laramie, Wyoming
- Lyon, Miss E. Lucille, Elmira College, Elmira, N. Y.
- MacDonald, James W., Chief, Civilian Graduate Branch, Graduate Section, A.A.F., Institute of Technology, Wright Field, Dayton, Ohio
- MacKenzie, Mrs. Barbara O'Neill, Personnel Associate, Brooklyn College, Brooklyn, N. Y.
- MacKenzie, Donald M., Registrar, Frances Shimer College, Mount Carroll, Illinois
- Madsen, Howard C., Manager, Technical Employment and Training, Westinghouse Electric and Manufacturing Company, 306 Fourth Avenue, East Pittsburgh, Pennsylvania
- Mallay, Miss Betty Anne, Director of Appointment Bureau, Manhattanville College of the Sacred Heart, New York, N. Y.
- Mallett, Donald R., Office of Student Affairs, Purdue University, West Lafayette, Indiana
- Manson, Dr. Grace E., 2200 Elsinor Avenue, Baltimore 16, Maryland
- Manzer, Dr. Charles W., Faculty Adviser, Washington Square College, New York University, New York, N. Y.
- Marks, Arlyn, Director of Nonacademic Personnel, State University of Iowa, Iowa City, Iowa
- Marshall, Dr. Thomas O., Dean of Students, American University, Washington, D. C.
- Mason, Dr. Harry M., Dept. of Psychology, Purdue University, Lafayette, Indiana
- Matthies, Roland Carl, Men's Adviser and Veterans Coordinator, Wittenberg College, Springfield, Ohio
- Maurer, Albert V., President, Concordia Teachers College, 7400 Augusta Street, River Forest, Illinois
- Meyer, Mrs. Ineva R., Assistant to the Dean of Arts and Science, 103 South Hall, University of Wisconsin, Madison, Wisconsin
- Miller, Miss Eleanor O., Acting Dean of Women, Illinois College, Jacksonville, Illinois
- Miller, Pres. Ernest Miller, Goshen College, Goshen, Indiana

- Miller, Prof. Ross., Religious Education, Jamestown College, Jamestown, North Dakota
- Mills, Miss Thelma, Director of Student Affairs for Women, University of Missouri, Columbia, Missouri
- Miner, Dr. Robert J., Miami University, Oxford, Ohio
- Mitchell, Dean William R., Dean and Registrar, Missouri Valley College, Marshall, Missouri
- Moore, Robert F., Secretary of Appointments, Alumni House, Columbia University, New York, N. Y.
- Morgan, E. Robins, Director of Placement, Lehigh University, Bethlehem, Pennsylvania
- Morgan, Miss Margaret, Head of Division of Teaching Appointments, Appointments Office, Columbia University, New York 27, N. Y.
- Morgan, Roy E., Director of Instruction, New Haven Y.M.C.A. Jr. College, 15 Prospect St., New Haven, Conn.
- Moritz, R. D., Director of Department of Educational Service, University of Nebraska, Lincoln, Nebraska
- Mueller, Dr. Kate Heyner, Dean of Women, Indiana University, Bloomington, Indiana
- Musgrove, Dr. Wilfred P., Director of Personnel, Huntington College, Huntington, Indiana
- McCall, Prof. William C., Department of Education, University of South Carolina, Columbia, South Carolina
- McCarn, Mrs. Ruth O., Counselor to Women, Northwestern University, Evanston, Illinois
- McCaul, Miss Margaret E., 4245 W. 51st Terrace, Mission, Kansas
- McClintock, Dr. James A., Personnel Office, Brothers College, Drew University, Madison, New Jersey
- McCracken, Charles William, Dean of Men, Muskingum, New Concord, Ohio
- McCreary, Dr. Otis C., Director of Training, Aluminum Company of America, Los Angeles, California
- McCully, C. Harold, Littleton, Colorado
- McCurley, Miss Mary T., Assistant to the President, Goucher College, Baltimore, Maryland
- McKinney, Prof. Fred, University of Missouri, Columbia, Missouri
- Newland, Lt. Col. T. Ernest, Associate Director, Military Psychology and Leadership, U. S. Military Academy, West Point, N. Y.
- Nichols, Dean Joseph C., 1983 E. 24th Street, Fenn College, Cleveland 15, Ohio
- Nield, Mrs. Marjory P., Director of Vocational Office, Smith College, Northampton, Mass.
- Nielsen, Dr. Otto Richard, Texas College of Arts and Industries, Kingsville, Texas
- Northrup, Miss Catherine, Assistant Director of Student Personnel, Colorado State College, Fort Collins, Colorado
- Norton, DeWitt C., Supervisor of Counseling and Testing, Cornell College, Mt. Vernon, Iowa



- Odom, Dr. Charles L., Veteran Administration Regional Office, Advisement and Guidance Section, 726 West End Blvd., New Orleans 19, Louisiana
- Olson, Miss Dorothy, Director of Texas Union, University of Texas, Austin 12, Texas
- Onthank, Dr. Karl W., Dean of Personnel Administration, Johnson Hall, University of Oregon, Eugene, Oregon
- Orr, Dean Cora I., Muskingum College, New Concord, Ohio
- Orr, James McK., Director of Student Personnel, New Haven Y.M.C.A. Junior College, 15 Prospect Street, New Haven 11, Connecticut
- O'Shea, Miss Harriet E., Division of Education and Applied Psychology, Purdue University, Lafayette, Indiana
- Owen, Mary C., Dean of Women, Scarritt College for Christian Workers, Nashville 4, Tennessee
- Palmerton, L. R., Director of Student Personnel, South Dakota School of Mines, Rapid City, South Dakota
- Parks, Donald S., Personnel Director, University of Toledo, Toledo, Ohio
- Partch, Dean C. E., School of Education, Rutgers University, New Brunswick, New Jersey
- Pattee, Howard H., General Secretary, California Association of Independent Secondary Schools, 645 W. 10th Street, Claremont, California
- Paul, Dr. J. B., Director of Bureau of Research, Iowa State Teachers College, Cedar Falls, Iowa
- Pearl, Miss Penelope M., Health Service, University of Michigan, Ann Arbor, Michigan
- Peck, Miss Margaret, Assistant Dean of Women and Counselor of Residences, University Residences for Women, University of Texas, Austin 12, Texas
- Peiffer, Herbert C., Dean of Student Personnel, San Diego State College, San Diego, California
- Penron, E. B., Oregon State College, Corvallis, Oregon
- Pepinsky, Prof. Harold B., Institute of Counseling, Testing and Guidance, Michigan State College, East Lansing, Michigan
- Perry, William D., North Carolina Mental Hygiene Society, Chapel Hill, North Carolina
- Pierson, George A., Dean of Students and Head of Department of Social Education, University of Utah, Salt Lake City, Utah
- Plank, Prof. William B., Department of Mining and Metallurgical Engineering, Lafayette College, Easton, Pennsylvania
- Polmantier, Paul C., Associate Professor of Education, University of Missouri, Columbia, Missouri
- Porter, Prof. J. F., 1357 E. 57th Street, Chicago, Illinois
- Price, Miss Louise, Counselor, Department of Personnel Service, Brooklyn College, Brooklyn, N. Y.
- Prince, Chester J., Personnel Director, William Jewell College, Liberty, Missouri

- Purdum, Dr. T. Luther, Director, Bureau of Appointments and Occupational Information, University of Michigan, Ann Arbor, Michigan
- Purdue University, Lafayette, Indiana
- Quick, Mariette, University of Pennsylvania, Philadelphia, Pennsylvania
- Ramsay, Miss L. Alice, Director of Personnel Bureau, Connecticut College, New London, Connecticut
- Read, Miss Julia E., Personnel Director, College of St. Elizabeth, Convent Station, New Jersey
- Reeves, Mrs. Mildred Jones, 363 Park Street, Hackensack, New Jersey
- Reilly, Miss Margaret M., Dean, Barat College of Sacred Heart, Lake Forest, Illinois
- Reinecke, Miss Virginia J., Assistant Counselor of Women, Faculty Exchange, University of Oklahoma, Norman, Oklahoma
- Replogle, Dr. Fred A., Stevenson, Jordan, Harrison, 205 W. Wacker Drive, Chicago, Illinois
- Reusser, Dr. Walter C., Director of Student Personnel and Guidance, University of Wyoming, Laramie, Wyoming
- Richardson, Thomas F., Director of Student Personnel, Texas Christian University, Fort Worth, Texas
- Riley, Sister John Marie, Director of Personnel, College of St. Teresa, 5700 Main Street, Kansas City 2, Missouri
- Riley, Miss Marguerite A., Placement Director, St. Joseph College, Asylum Avenue, West Hartford, Connecticut
- Robison, Miss Sallie E., Personnel Director, Louisiana Polytechnic Institute, Ruston, Louisiana
- Rogers, Miss Margaret, Director of Appointment Bureau, Manhattanville College of Sacred Heart, Convent Ave., 133rd St. W., New York 27, N. Y.
- Rohrer, John H., Assistant Professor of Psychology, University of Oklahoma, Norman, Oklahoma
- Roy, Howard L., 1117 4th Street, N.E., Rochester, Minnesota
- Ruedisili, Dean C. H., 103 South Hall, University of Wisconsin, Madison 6, Wisconsin
- Ruhlman, Miss Jessie L., Dean of Women, 405 Hilgarde Ave., Los Angeles 24, Calif.
- Russell, Dr. John Dale, North Central Association, Commission on Colleges and Universities, 5835 Kimbark Ave., Chicago 37, Illinois
- Sackett, Dr. Everett B., Registrar Director of Admissions, University of New Hampshire, Durham, New Hampshire
- Sage, Nathaniel M., Placement Officer, Massachusetts Institute of Technology, Cambridge, Massachusetts
- Samuell, James M., Dean of Men, Muskingum College, New Concord, Ohio
- Saul, Herbert W., 18 Talbot Road, Braintree, Massachusetts
- Schaper, Dr. Florence W., Director of Student Guidance and Personnel, Lindenwood College, St. Charles, Missouri

- Schmidt, Carl H., Director of Personnel, N.D.A.C. Station, Fargo, North Dakota
- Schneiders, Dr. Alexander A., Director of Student Personnel, Loyola University, Chicago, Illinois
- Schneidler, Dr. Gwendolen G., Vocational Advisement Supervisor, Veterans Administration, Munitions Building, Washington 25, D. C.
- Schonfeld, Dr. Louise Rost, 211 W. 106th Street, New York 25, N. Y.
- Seaman, William H., Director of Admissions and Bureau of Appointments, Oberlin College, Oberlin, Ohio
- Seamans, Harry W., 2810 Naylor Road, S.E., Apt. A137, Washington, D. C.
- Seawell, Ann, Director of Placement, University of Georgia, Athens, Georgia
- Seibert, Earl W., Juniata College, Huntingdon, Pennsylvania
- Sellen, A. G., Dean of the College, Washburn Municipal University, Topeka, Kansas
- Sellers, Miss Ava Foster, Personnel Assistant to Dean of Women, Vanderbilt University, Nashville 4, Tennessee
- Shaffer, Robert H., Assistant to Dean, Junior Division, Indiana University, Bloomington, Indiana
- Shank, Dr. Donald J., School of Industrial Relations, Cornell University, Ithaca, N. Y.
- Shell, Miss Helen H., 70 Hawthorne Place, Upper Montclair, New Jersey
- Showalter, Dean D. F., Arkansas State College, Jonesboro, Arkansas
- Shuman, Helen A., Dean of Women, Southern Illinois Normal University, Carbondale, Illinois
- Shybekay, Dr. Derso, Director of Market Research, North Star Woolen Mill Company, Minneapolis, Minnesota
- Sisson, Lt. Donald E., Classification and Replacement, 270 Madison Avenue, New York, N. Y.
- Sistrunk, Henry L., Y.M.C.A., 19 LaSalle Street, Chicago, Illinois
- Skinner, Dr. Laila, Dean of Women, Allegheny College, Meadville, Pennsylvania
- Smiley, E. Kenneth, Director of Admissions, Lehigh University, Bethlehem, Pennsylvania
- Smith, Dr. Margaret Ruth, Associate Admissions Officer, Wayne University, Detroit, Michigan
- Smoke, Dr. Kenneth L., 249 N. Washington St., Gettysburg, Pa.
- Smutz, Harold T., Chief of Regional Examining and Placement Division, 9th U. S. Civil Service Region, 635 New Federal Building, St. Louis 1, Missouri
- Spacie, Edwin G., Student Counselor, Michigan College of Mining and Technology, Houghton, Michigan
- Spaulding, Miss Helen, Vocation Counselor and Assistant, Colby Junior College, New London, New Hampshire
- Spindt, Herman, Bureau of Guidance and Placement, University of California, Berkeley, California

- Stalnaker, Armand C., Placement Manager, College of Commerce, Ohio State University, Columbus, Ohio
- Stapleton, Miss Mary R., Associate in Personnel Bureau, Brooklyn College, Bedford Avenue and Avenue H, Brooklyn 10, N. Y.
- St. Clair, Walter F., 25 Waldo Street, Manchester, New Hampshire
- Stedman, Miss Edith G., Director of Appointment Bureau, Radcliffe College, Cambridge, Massachusetts
- Stephanus, Miss Jane E., Acting Placement Director, University of Detroit, McNichols Road, Detroit 21, Michigan
- Stevens, Prof. Edwin B., College of Education, University of Washington, Seattle, Washington
- Stogdill, Dr. Emily L., Department of Psychology, Ohio State University, Columbus, Ohio
- Stone, H. E., Secretary, Loans and Placement, West Virginia University, Morgantown, West Virginia
- Strang, Prof. Ruth, Records and Research, College of Education, Teachers College, Columbia University, New York, N. Y.
- Stuit, Dr. Dewey B., Department of Psychology, University of Iowa, Iowa City, Iowa
- Super, Dr. Donald E., Teachers College, Columbia University, New York 27, N. Y.
- Sutherland, Robert L., University of Texas, Austin 12, Texas
- Swann, Prof. Reginald L., Teachers College of Connecticut, New Britain, Connecticut
- Swanson, Prof. Donald E., Director of Student Personnel, Hamline University, St. Paul, Minnesota
- Swenson, Miss Hilda G., Social Director, Box 151, College Station, Washington State College, Pullman, Washington
- Tallmadge, Miss Frances M., 328 W. 88th Street, New York 24, N. Y.
- Tarbox, Sidney, Director of Recruitment of 7th Regiment of U. S. Civil Service Commission, 12 S. Monroe Street, Hinsdale, Illinois
- Taylor, Mary E., Women's Residence Center, University of Indiana, Bloomington, Indiana
- Thisted, M. N., Dean of Men, Western Illinois State Teachers College, Macomb, Illinois
- Thompson, Albert S., Director of University Counseling Service, Vanderbilt University, Nashville 4, Tennessee
- Thompson, C. Woody, Director of Office of Student Affairs, University of Iowa, Iowa City, Iowa
- Thompson, Ralph D., Executive Secretary to the President, University of Utah, Salt Lake City 1, Utah
- Thornbury, Miss Zita L., Director of Vocational Bureau, Vassar College, Poughkeepsie, New York
- Thorpe, Miss Alice L., Appointment Office, Wheaton College, Norton, Massachusetts
- Tibbals, Dean C. A., Illinois Institute of Technology, Chicago, Illinois
- Tibbetts, Clark, Director of Institute of Human Adjustment, University of Michigan, Ann Arbor, Michigan
- Tiner, Pres. Hugh M., George Pepperdine College, 1121 W. 79th Street, Los Angeles, California

- Todd, J. Edward, 822 E. John Street, Appleton, Wisconsin
- Todd, Margaret E., Assistant Director, Social Education, Residence Halls for Women, Purdue University, Lafayette, Indiana
- Trabue, Dean M. R., School of Education, Penn State College, State College, Pennsylvania
- Triggs, Dr. Frances, Kingscote Apartment 3G, 419 W. 119th Street, New York 27, N. Y.
- Turner, Mrs. Lillian Barbour, Shurtleff College, Alton, Illinois
- Twyman, Mrs. Margaret G., Director of Placement Service, University of Oklahoma, Norman, Oklahoma
- Tyler, Miss Leona E., Director of Bureau of Personnel Research, University of Oregon, Eugene, Oregon
- Valley, John R., Acting Head, Veterans Administration Guidance Center, Personnel Research Institute, Western Reserve University, 1801 Stanwood Rd., Suite 7, East Cleveland 12, Ohio
- Van Dusen, Dr. Albert C., Director of University College Bureau of Counseling, Department of Psychology, Northwestern University, Evanston, Illinois
- Van Wagenen, Dr. (Miss) Beulah C., Girl Scouts, Inc., 155 East 44th Street, New York, N. Y.
- Voorhees, Dean Edwin C., University of California, Berkeley, California
- Voorhees, Miss Helen M., Director of Appointment Bureau, Mount Holyoke College, South Hadley, Massachusetts
- Waldrop, Robert S., 1618 South Boulevard, Ann Arbor, Michigan
- Wallace, Mrs. Frances Gibson, Freshman Adviser, 1520 Fifth Ave., Huntington, W. Virginia
- Wallace, Dr. Isabel King, Vocational Counselor for Women, University of Rochester, Rochester, N. Y.
- Walter, Edwin Carlton, Director of Student Personnel, Pacific College, Angwin, Calif.
- Walters, Dr. Charles E., Pres., Alfred University, Alfred, N. Y.
- Wangemann, Dr. Charles E., Head of Bureau of Placement, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
- Wegener, Miss Mary A., Appointments Office, Columbia University, New York, N. Y.
- Weigel, Virginia, Dean of Students, Frances Shimer College, Mt. Carroll, Ill.
- Weir, Miss Edith, Director of Bureau of Teacher Placement, University of Southern California, Los Angeles, California
- Wells, Charles Arthur, Professor of Psychology, American International College, Springfield, Massachusetts
- Wengate, Miss Pauline, 123½ Glessner Street, Americus, Georgia
- Wessell, Nils Y., Dean of Men and Director of Admissions, Tufts College, Medford 55, Mass.
- Weston, Miss Janet L., 112 Commerce Building, University of Illinois, Urbana, Illinois
- Westover, Mrs. Ada S., 1900 Lyons Street, Lincoln, Nebraska
- Westover, Dr. Frederick L., Student Personnel Assistant to the Dean,

- Brooklyn College, Bedford Avenue and Avenue H, Brooklyn, N. Y.
- Whiteman, Dr. Harriet E., Director of Student Personnel, New Jersey State Teachers College, Newark, New Jersey
- Wightwick, Miss M. Irene, Personnel Director, College of New Rochelle, New Rochelle, New York
- Wilcox, Prof. E. R., College of Engineering, University of Washington, Seattle, Washington
- Williams, Cornelia DeCamp, General College, University of Minnesota, Minneapolis, Minnesota
- Williams, Prof. Lewis W., Secretary, Appointment Committee, College of Education, University of Illinois, Urbana, Illinois
- Williamson, Prof. E. G., Dean of Students, University of Minnesota, Minneapolis, Minnesota
- Wilson, Eugene S., Amherst College, Amherst, Massachusetts
- Winegar, Mrs. R. M., Box 188, Aptos, California
- Winningham, Miss Jane Louise, Huron Valley Children's Center, 104 Welch Hall, Ypsilanti, Michigan
- Winslow, Miss Mary Isabel, Fonthonne College, Wydown and Big Ben Bldgs., St. Louis, Missouri
- Woellner, Prof. Robert C., Secretary, Board of Vocational Guidance and Placement, University of Chicago, Chicago, Illinois
- Wood, Austin, Assistant Professor of Psychology, Brooklyn College, Brooklyn, N. Y.
- Woodhouse, Mrs. Chase Going, Director of Institute of Women's Professional Relations, Connecticut College, New London, Connecticut
- Wrenn, Dr. C. Gilbert, University of Minnesota, Minneapolis, Minnesota
- Zerfoss, Prof. K. P., Department of Psychology, George Williams College, 5315 Drexell Avenue, Chicago, Illinois

## THE CONTRIBUTORS

**William D. Altus**—Ph.D., New York University, 1941. Instructor, Santa Barbara State College, 1941–1944; Assistant Professor, 1944–, Santa Barbara College, University of California. On military leave, as Psychologist with the U. S. Armed Forces, 1942–1946. Fellow, American Psychological Association.

**Hugh M. Bell**—Ph.D., Stanford University, 1941. Dean of Student Personnel and Guidance and Professor of Psychology, Chico State College. Acting Associate Professor of Psychology, University of Minnesota, summer 1938. With U. S. Armed Forces, 1942–1945. Psychologist with the Veterans Administration, Washington, D. C., 1945–1946. Author of articles, monographs, and tests. Member, American Psychological Association, American Association for Applied Psychology, American College Personnel Association, National Vocational Guidance Association, American Educational Research Association.

**Robert Callis**—M.A., University of Minnesota, 1946. With the U. S. Navy, 1942–1946. Counselor, General College, and graduate student, University of Minnesota, 1946–.

**M. Catherine Evans**—Ph.D., University of Minnesota, 1940. Research Assistant, University of Minnesota, 1938–1939. Assistant to the Director of Evaluation Study of Bennington College, 1939–1941. Assistant Professor and Vocational Counselor, University of Indiana, 1941–. Author of articles on personality, measurement, and counseling. Associate Member, American Psychological Association. Member, American College Personnel Association, National Association of Deans of Women, and National Vocational Guidance Association.

**Daniel D. Feder**—Ph.D., University of Iowa. Associate, University of Iowa, 1934–1938. Assistant Director, Personnel Bureau, and Assistant Professor of Psychology, University of Illinois, 1938–1942. Executive Officer and Supervisor, Illinois State Civil Service Commission, 1942–1946. Dean of Students and Professor of Psychology, University of Denver, 1946–. On military leave for service with the U. S. Navy, 1942–1946. Officer in Charge, Radio Materiel Unit (formerly Training Activity, now part of Research Activity). Officer in charge to study German Naval Selection and Training Methods

attached to the U. S. Naval Technical Mission in Europe. Author of articles on personnel work and measurement. Member, American Educational Research Association, American Psychological Association, American College Personnel Association, Civil Service Assembly. President, American College Personnel Association, 1946-1947.

**Warren K. Layton**—Ph.D., University of Michigan. Three years in employment with a manufacturing concern (advertising and sales promotion); two years in the United States Army (psychological examining and personnel work); 30 years in various educational positions, including 25 years in the Detroit Public Schools, the last 15 as Director of Guidance and Placement. Special lecturer or summer session instructor at various times, the University of Illinois, the Michigan State Normal College, Wayne University, and Harvard University. Author of articles to several journals including *School and Society*, *Occupations*, *Education*, *School Review*, and the *Yearbooks* of the National Society for the Study of Education. Community activities have included service in the Boy Scouts, the Y.M.C.A., the Board of Directors and the Case Committee of the Children's Aid Society, the Detroit Council of Veterans Affairs, the Detroit Council for Youth Service, the Wayne County Youth Farm Labor Committee, the National Vocational Guidance Association (Vice-President) and the Council of Guidance and Personnel Associations (Past President).

**Harold B. Pepinsky**—Ph.D., University of Minnesota, 1946. Teacher and Research Assistant, 1938-1942; Instructor, College of Education, summer 1943; Administrative Fellow, Dean of Students' Office, 1942-1944, University of Minnesota. Assistant Director, Guidance Bureau, and Assistant Professor of Education, University of Kansas, 1944-1945. Assistant Professor, Institute of Counseling, Testing and Guidance; Assistant Professor, Department of Effective Living, Michigan State College, 1946-. Author of articles on counseling and co-author of *Financial Assistance for College Students*. Associate Member, American Psychological Association. Member, American College Personnel Association, National Vocational Guidance Association, Phi Delta Kappa.

**Gwendolen Schneidler**—Ph.D., University of Minnesota, 1940. Research Assistant, Employment Stabilization Research Institute, University of Minnesota and Minnesota State Employment Service, 1931-1933. Technical Assistant, Adjustment Service, New York, 1933-1934. Teaching Assistant in Psychology, University of Minnesota, 1934-1937. Instructor, Goucher College, 1937-1939. Counselor, Student Counseling Bureau, University of Minnesota, 1939-1942. Lt. (U.S.N.R.), Assistant Officer-in-Charge, Enlisted Selection, Bureau of Naval Personnel, 1942-1945. Chief, Educational Counseling Division, Veterans Administration 1945-. Co-author of *Student Guidance Techniques* and author of articles in educational



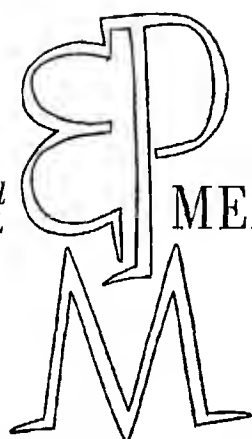
and psychological journals. Member, American Psychological Association, American College Personnel Association, National Vocational Guidance Association.

Edmund G. Williamson—Ph.D., University of Minnesota, 1931. Assistant Professor of Psychology, 1930–1938; Associate Professor of Psychology, 1938–; Director of University Testing Bureau, 1932–1938; Coordinator of Student Personnel Services, 1938–1941; Dean of Students and Professor of Psychology, 1941–, University of Minnesota. Author of books and articles on counseling and personnel work. Member, American Psychological Association, Sigma Xi, Phi Delta Kappa, Psi Chi, Alpha Kappa Lambda, American College Personnel Association, American Association for Applied Psychology, Advisory Committee to the United States Armed Forces Institute Editorial Board, EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT Chairman, Committee on Student Personnel Work of American Council on Education. Recently returned from three-month survey of industrial and educational personnel work in Germany.

C. Gilbert Wrenn—Ph.D., Stanford University, 1932. Vocational Counselor, Stanford University, 1928–1936. Associate Director, General College; Associate Professor of Educational Psychology, 1936–1938; Professor of Educational Psychology, 1938–, University of Minnesota. Consultant, Student Personnel Teacher Education Commission of the American Council on Education, 1939–1942. On military leave with the U. S. Armed Forces, Personnel Officer in Bureau of Naval Personnel and Pacific Area, 1942–1946. Associate, American Youth Commission, 1939–1941. Author and co-author of *Student Personnel Problems*, *Studying Effectively*, *Aids to Group Guidance*, *Time on Their Hands*, and numerous journal articles. President, National Vocational Guidance Association. Vice-President, Council of Guidance and Personnel Association, 1946–.



EDUCATIONAL and  
PSYCHOLOGICAL



MEASUREMENT

**VOLUME SEVEN, NUMBER TWO, SUMMER**

<i>A Critical Examination of the Concepts of Face Validity.</i>	
CHARLES I. MOSIER .....	191
<i>Test Patterns in the Vocational Clinic.</i>	LINDSEY R. HARMON
	207
<i>The Ergic Theory of Attitude and Sentiment Measurement.</i>	
RAYMOND B. CATTELL .....	221
<i>The Cornell Technique for Scale and Intensity Analysis.</i>	
LOUIS GUTTMAN .....	247
<i>An Empirical Study of the Relationship between Item Validity and Internal Consistency.</i>	WILLIAM A. OWENS ....
	281
<i>A New Approach to Oral Testing.</i>	WILLIAM BRODY and
NORMAN J. POWELL .....	289
<i>The Values of an Educational Guidance Clinic.</i>	JOSEPH C. HESTON .
	299
<i>The Number Relations Section of the Carnegie Mental Ability Tests Treated as a Power Test.</i>	WILEY S. GARRETT ..
	309
<i>Prediction of the Difficulty Index of Objective-Type Spelling Items.</i>	C. W. HARRIS .....
	319
<i>Studies of Popularity in College: II. Do Dormitory Arrangements Affect Popularity?</i>	JEAN WAID REILLY and FRANCIS P. ROBINSON .....
	327
<i>A Study of the Effect of Conformity to Social Expectancy on Evaluative Attitudes.</i>	ROBERT P. FISCHER and AYVONNE L. ANDREWS .....
	331
<i>Measurement Abstracts</i> .....	337
<i>The Contributors</i> .....	351

PRINTED IN THE UNITED STATES OF AMERICA  
THE SCIENCE PRESS PRINTING COMPANY  
LANCASTER, PENNSYLVANIA

# A CRITICAL EXAMINATION OF THE CONCEPTS OF FACE VALIDITY

CHARLES I MOSIER

Office of the Secretary of War<sup>1</sup>

FACE validity is a term that is bandied about in the field of test construction until it seems about to become a part of accepted terminology. The frequency of its use and the emotional reaction which it arouses—ranging almost from contempt to highest approbation—make it desirable to examine its meaning more closely. When a single term variously conveys high praise or strong condemnation, one suspects either ambiguity of meaning or contradictory postulates among those using the term. The tendency has been, I believe, to assume unaccepted premises rather than ambiguity, and beautiful friendships have been jeopardized when a chance remark about face validity has classed the speaker among the infidels.

An examination of the ways in which the term "face validity" has been used indicates three frequent meanings. These are sufficiently similar as to be confused, yet so different in their implications that to understand one meaning where another was intended leads to a wholly erroneous interpretation. This paper will analyze the various meanings which have been attributed to the term and it will then recommend that the term (and one of its meanings as well) be banished to outer darkness.

The three meanings which have been attributed to the term may be characterized as: (1) validity by *assumption*, (2) validity by *definition*, and (3) the *appearance* as well as the reality of validity. A fourth concept, validity by hypothesis, is closely related to the first two and deserves consideration in

---

<sup>1</sup> Opinions expressed in this paper are those of the author and do not necessarily reflect the policies of the War Department

connection with them, although this concept has not generally been termed "face validity."

*Validity by assumption:* As used in this way, the term "face validity" carries the clear meaning that a test is assumed to be valid for the prediction of an external criterion if the items which compose it "appear on their face" to bear a common-sense relationship to the objective of the test. The assumption of validity in this case is asserted to be so strong that statistical evidence of validity is unnecessary; indeed, statistical evidence showing a lack of validity may be set aside by the strength of the assumption.

*Validity by definition:* For some tests, the objective is defined solely in terms of the population of questions from which the sample comprising the test was drawn, e.g., when the ability to handle the one hundred number facts of addition is tested by a sampling of those number facts. In these cases, the test is considered to be valid if the sample of items appears to the subject-matter expert to represent adequately the total universe of appropriate questions. The objective of the test is so defined that the index of reliability (the square root of the reliability coefficient) is, by definition, the measure of validity. This is so, not because of a definition of validity, but because of the way the objective of the test is defined. This situation is the one for which the term "face validity" was apparently coined.

*Appearance of validity:* In this usage, the term "face validity" implies that a test which is to be used in a practical situation should, in addition to having pragmatic or statistical validity, appear practical, pertinent and related to the purpose of the test as well; i.e., it should not only *be* valid but it should *also appear* valid. This usage of the term assumes that "face validity" is not validity in any usual sense of the word but merely an additional attribute of the test which is highly desirable in certain situations.

*Validity by hypothesis:* This concept, not generally associated with the term "face validity," is nevertheless sufficiently related to validity by assumption and validity by definition as to call for analysis at this point. The term "validity by hy-

pothesis" is used to characterize the following situation. Often, before the validity of a test can be empirically verified for a particular group by demonstration of its relationship to a satisfactory criterion, the test must be used to meet an immediate practical need. In such instances, the use of the test involves the hypothesis that it has a useful degree of validity. This hypothesis is based upon the designed similarity of the particular test to other tests already demonstrated to have known validity for the purpose in question. The validity of the test is not assumed in the sense that no further proof is required; neither is the objective of the test defined in such a way that the reliability of the test is evidence of its validity for the defined purpose. Rather the hypothesis is stated that, because of the sum total of previous knowledge relating to methods of predicting this particular criterion it is reasonable to suppose that a test of this sort will prove to be valid by the conventional statistical tests. This reasonable presumption, however, is subject to empirical verification by fact. Pending the opportunity for such verification, the presumption may be sufficiently strong as to justify the use of the test. Similarly, the physician studies the symptoms and the general condition of the patient and then, on the basis of his knowledge of the past effects of remedies upon similar symptoms in similar patients, prescribes treatment. He does this even though this combination of remedies has not occurred before in his experience and certainly not with this patient (who may have an unsuspected allergy which will defeat the purpose of the remedy).

With these four possible meanings of the term before us, it becomes profitable to examine each one in more detail.

### *Validity by Assumption*

This conception of "face validity" is illustrated by the following quotations from a widely circulated testing handbook:

Generally speaking, the validity of the test is best determined by using common sense in discovering that the test measures component abilities which exist both in the test situation and on the job. This common-sense approach to the problem of validity can be strengthened greatly by basing the estimate of the component of the job on a systematic observation of job analysis.

The term "face validity" is thus used to imply that the appearance of a relationship between the test and the external criterion is sufficient evidence of pragmatic validity. This use is a pernicious fallacy. This illegitimate usage has cast sufficient opprobrium on the term as to justify completely the recommendation that it be purged from the test technicians' vocabulary, even for its legitimate usage. The concept is the more dangerous because it is glib and comforting to those whose lack of time, resources, or competence prevent them from demonstrating validity (or invalidity) by any other method. Moreover, it is readily acceptable to the ordinary users of tests and its acceptance in these quarters lends the concept strength. This notion is also gratifying to the ego of the unwary test constructor. It implies that his knowledge and skill in the area of test construction are so great that he can merrily design a test with the desired degree of effectiveness in predicting job success or in evaluating defined personality characteristics, and that he can do this so accurately that any further empirical verification is unnecessary. So strong is this ego complex that if statistical verification is sought and found lacking, the data represent something to be explained away by appeal to sampling errors or other convenient rationalization, rather than by scientific evidence which must be admitted into full consideration.

The concept of validity by assumption gains strength from the legitimate use of the term "face validity" to mean validity by definition. The superficial similarity, however, between the two concepts should not deceive us into accepting either the truth of the one or the necessary falsity of the other.

Any experienced test constructor can cite numerous instances of tests which appear so closely related to the external criterion that a high validity coefficient seems inevitable. The following example is to be considered merely one illustration which most readers can reproduce almost without limit from their own experience.

Two test construction agencies, each having a fairly large and competent staff, began work about the same time on an objective test to measure the clerical skills involved in alpha-



betical filing. Up to a certain point the two agencies worked independently, each devising its own test. Agency A, after an analysis of the job, constructed a test of which the following item is representative:

"Below are five names, in random order. If the names were placed in strict alphabetical order, which name would be *third*: (1) John Meeder; (2) James Medway; (3) Thomas Madow; (4) Catherine Meagan; (5) Eleanor Meehan."

The second agency designed a test of skill in alphabetical filing in which the task was as follows:

"In the following items you have one name which is underlined and four other names in alphabetical order. If you were to put the underlined name into the alphabetical series, indicate by the appropriate letter where it would go:

Robert Carstens

A. \_\_\_\_\_

Richard Carreton

B. \_\_\_\_\_

Roland Casstar

C. \_\_\_\_\_

Jack Corson

D. \_\_\_\_\_

Edward Cranston

E. \_\_\_\_\_

There was a general agreement that each of these tests was face-valid and that each consisted of work-samples representative of the filing of alphabetical material. It was also agreed that if one were going to use two different tests to measure filing ability, it would be difficult to get two tests more closely similar than these and still have different tests. Had the concept of validity by assumption prevailed, there is little question that each test would have been considered highly valid.

An actual tryout, however, revealed quite different results from those expected. The correlation of the two tests in a sample of 43 clerical workers was .01, although the Kuder-Richardson reliabilities of the two tests were .81 and .89, respectively. We have here two tests which, on the basis of face validity by assumption, would be equally valid but which correlate substantially zero with one another. If one is valid, the other is not likely to be. What happens when the two tests

are studied, not for their correlation with each other, but for their correlation with what seems to be a reasonable criterion, namely supervisors' ratings of speed and accuracy in filing? For 72 employed workers where accuracy of filing materials was an important part of the job, the correlation between the first of the two tests described and the supervisors' rating was .09.<sup>2</sup> For the second test the correlation with the supervisors' ratings of accuracy in alphabetizing was .00. (That these results cannot be attributed to the unreliability of the supervisors' ratings is indicated by correlation coefficients of .40 and above between the same ratings and scores on other tests.) These two examples, therefore, as well as those which the reader's experience will readily bring to bear, are sufficient to demonstrate the fallacy involved in the statement that a test can be assumed to be valid without further verification if only it "measures component abilities which are judged by common sense to exist both in it and in the job."

#### *Validity by Definition*

The foregoing discussion has assumed an outside criterion measurable apart from the test itself. The discussion which follows is applicable rather to the situation, very frequent in educational measurement, in which the only available measure of the criterion (that which the test is intended to measure) is, because of the nature of the criterion, directly and intimately related to the test questions themselves. If the objective is the measurement of the pupils' skill in forming the elementary number combinations of addition, a test consisting of the one hundred possible combinations is presumably valid by definition. In this case the index of reliability can be taken as the validity coefficient. Even in this simple situation, the actual validity is limited by the reliability of the particular test, by the form in which the problems are presented, e.g., in words, in columns or in equations (e.g., four plus two equals —;  $\overset{4}{+} 2$ ;  $4 + 2 = \text{—}$ ), the arrangement of the items and by the conditions of administration. As soon, however, as the test is

<sup>2</sup> The test did, however, show substantial correlation with other clerical skills and hence was useful in a general clerical battery, though not for its "face-valid" objective.

reduced from the totality of all situations which constitute the objective of measurement to a sample of those situations, the question recurs as to the extent to which the universe can be predicted from the sample. Moreover, it must be remembered that the relationship between test items and criterion behavior requires careful scrutiny. It is quite possible to design a test which apparently depends on the ability to perform the indicated additions, but is at the same time so dependent on verbal facility in understanding the directions, on speed of reaction, and on coding skills needed to record the answers, that the similarity between test situation and criterion situation is more apparent than real.

A further point which must be remembered in interpreting validity by definition is that it is frequently possible to establish several definitions of the criterion behavior, each obviously valid and yet each bearing far less than perfect relationship to the other. In the investigation of spelling ability, one obviously valid criterion of ability to spell might be the number of words correctly spelled from dictation. Should the words be dictated singly or in sentences, in a Brooklyn, Mobile, or Chicago accent? Another criterion which might be used, however, is a count of the number of words misspelled in compositions written by the pupils. Either of these criteria is, upon its face, a valid reflection of spelling ability. Nevertheless, empirical investigation is unlikely to show a perfect correlation between dictation and correct spelling in compositions, even after correction for attenuation. Which universe should be sampled to provide a face-valid test of spelling?

Finally, in the validation of a test by definition, it must be remembered that *the direction of the argument flows from the test to the definition of the criterion* rather than from the conceptually defined criterion to the test as a valid measure. The only proper statement which can be made about a test in terms of face validity by definition is that this test is a valid measure of that and only that universe of individual behavior patterns for which these items constitute a representative sample. If one is prepared to infer such a universe and consider *that* universe rather than one defined in any other way, such a concept

of validity may be useful. The necessity for inferring the conceptual nature of the universe from an examination of the sample still exists as a judgmental process and as one which is peculiarly subject to error.

If we return to the example of the two alphabetizing tests given in the section above we see how readily one may be misled into generalizing beyond the nature of the facts given. It is not difficult to draw the conclusion, from an inspection of the items, that these two tests were representative of the same universe and that therefore either test is a valid measure of the same set of skills. The fallacy of the conclusion, however, is attested by the absence of correlation between the two tests as cited above.

In educational achievement tests it is possible to outline the concepts to be covered in a particular course of study. These concepts may be sampled so systematically and so comprehensively that we are prepared to say the test questions constitute an adequate representation of all of the questions which might be asked on this course, in the light of its content and stated objectives. Even so, the questions may be so formulated that the crucial skills for achieving a high score on the examination are quite different from a knowledge of course content and the achievement of the stated objectives. We are correct in saying that the test is a valid measure of "whatever it measures reliably." We may be far from correct in inferring that the hypostatized "whatever" is what it appears to be on the face of the test. Nevertheless if we rely on validity by definition, we face the obligation of defining that "whatever" in some meaningful terms without running into the pitfall of *assuming* that the "whatever" is synonymous with the test constructor's objective in preparing the test.

As we examine critically the distinction between validity by assumption and validity by definition, we are led to see how tenuous is the dividing line between the scientifically defensible use, "validity by definition," and the totally unscientific and indefensible use, "validity by assumption."

Moreover, we do not escape the dilemma by refusing to recognize anything except external criteria. The validity of

the external criterion is just as much open to question as is the validity of the test which is being checked against it. Consider the situation in which a test purporting to measure clerical aptitude is "validated" by correlating test scores with salary (where salary is presumed to reflect the level of duties and responsibilities assigned). A high correlation between test score and salary level might well be taken, however, not as validation of the test but as validation of the agency's promotional system and an indication of the effectiveness with which the placement office had sought out and recommended for promotion the employees with the highest level of knowledge and skill. As Toops has pointed out, the criterion is a complex and elusive concept.<sup>3</sup> This paper is not the place for a systematic analysis of the nature of the criterion. It suffices to point out here that it is frequently possible to define in verbal, as distinct from operational, terms a criterion which is a socially significant independent measure of the behavior to be predicted by the test; such a definition is not in itself a sufficient guarantee that the criterion used to validate the test is itself valid.

### *The Appearance of Validity*

In many situations it is highly desirable that the testing instrument should have a high degree of "consumer acceptance." These situations are most commonly found in, but by no means limited to, the field of employment testing. If a test is to be used effectively in achieving its objectives, it is essential that it actually be selected for use and that the results of the test be acceptable to those responsible for action on the basis of these results. In the area of public employment testing, e.g., civil service examining, the test must be acceptable not only to those using the test but to those taking the test as well. To a large extent this is also true in educational situations, particularly in the field of counseling. Up to a certain point the acceptability of the test can be carried by weight of authority. The board of examiners, the test technicians, or the counseling experts assert on the basis of their technical knowledge that the test is good, and their assertion is accepted without question.

<sup>3</sup> Toops, H. A. "The Criterion." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, IV (1944), 271-297.

In other situations, however, this assertion of authority is not sufficient to carry conviction. Moreover, the technical evidence on which such authoritative statements should be based is often neither comprehensible nor completely convincing to those who must be convinced.

In Civil Service situations, the candidate whose score is less than he expected is inclined to attribute his low score, not to his own deficiencies but to the impractical nature of the test in relation to the job for which he is being examined. His dissatisfaction with the test results and his feeling of injustice may, of course, have real merit. We have not yet reached the era of public personnel examining where all tests are technically sound. Whether or not there is merit in his claim, the legislature, the courts, and public opinion, the court of last appeal, are more readily impressed by superficial appearances than by correlation coefficients. It becomes highly important, therefore, that a test to be used in such a situation not only *be* valid in the pragmatic sense of affording reasonably accurate predictions of job competence, but *have the appearance of validity* as well.

This appearance of validity as an added attribute is important in terms of the acceptance of the test, not only by the persons being examined, but also by those operating officials who are charged with the responsibility for taking action based upon the test results. If sound tests are given and accurately reported, but the supervisor, interviewer, or counselor has no confidence in them, the results will not be used effectively.

In passing it should be noted that the concern of the Civil Service or merit system agency with the consumer acceptance of the test should not be merely a negative one of avoiding appeals or legislative pressures. In a democratic society the quality of public service is dependent to a large extent upon the public's opinion of the quality of public servants. If the examination by which public servants are selected (whether it be an objective test or an examination of the candidates' voting records) is such that competent persons in a particular occupation are convinced that they have no opportunity to demonstrate their competence, they will not file for the examination.

or apply for the position. Since even the best Civil Service system can do no more than to select the best qualified persons of those who apply for positions, it is essential that every possible step be taken to insure that the most competent ones make application. They certainly will not do so if they believe that their examination will be impractical, theoretical, and deny them an opportunity to demonstrate their real ability. Moreover, in the face of such an attitude, statistical evidence on the validity of the test is likely to prove convincing only after an educational campaign extending to several generations of test-takers.

The foregoing discussion does not imply that predictive value is to be sacrificed to superficial appearances. Neither does it imply that a statistically valid test may be used only if it also has the appearance of practicality. It does imply, however, that the appearance of practicality is an objective sufficiently desirable in its own right that it may often be sought as an additional end consistent with the principal objective—predictive value.

The use of the term "face validity" to denote the appearance of a relationship to job performance as an attribute in addition to rather than instead of a statistical relationship, is frequently and unjustifiably confused with the notion of "face validity" by assumption. There is, however, a much clearer distinction between these two usages than between validity by assumption and validity by definition.

#### *Validity by Hypothesis*

This fourth view of validity has not, to the writer's knowledge, been explicitly termed face validity, although it contains certain elements of confusion with validity by assumption. In the construction of any test it is necessary to formulate certain hypotheses as to the most valid type and content to achieve a particular purpose. These hypotheses are held with a greater or less degree of confidence depending upon (a) the amount and the convincingness of available data showing that test items X have proved valid in situation Y, (b) the similarity of test item X to the proposed test items X', and (c) the degree

of similarity between situation Y and situation Y' in which the test is to be used. If the new test is very similar to one previously shown to be valid and if the new situation is very similar to that in which the test was valid, then we may proceed with a high degree of confidence that the proposed test will be valid in the situation in which it is to be used. This confidence, of course, never approaches certainty, and a verification of the hypothesis is always necessary.

Even though the questions and the methods of administration are identical for the two tests (if we may speak of two sets of identical questions as two tests), the measuring instrument will not be identical in its effect if its application has shifted from one group of subjects to another or from one testing situation to another. When a test has been adequately standardized on one population and found to be highly valid for the prediction of a particular skill in that population, the use of the same test for another population involves merely a hypothesis, rather than the certainty, of its validity as a measure of the same skill in the new situation. Even though we may have a high degree of confidence that the hypothesis will be confirmed, it is nevertheless a hypothesis. As we construct alternate forms of a test and apply them to new situations to predict the same set of skills, our degree of confidence becomes substantially less. The confidence level is also reduced when we use the same test to predict a somewhat different set of skills. For example, a test may be used to predict competence in clerical office work of a certain type in one agency when the test has been validated against proficiency in office work of a similar type but in another agency. In all these cases we are dealing with varying amounts of confidence in the validity of a test in a particular situation. The degree of confidence which justifies the use of an examining instrument in advance of its validation in the specific situation is a question of administrative judgment which is not wholly answerable by statistical techniques.

The foregoing discussion makes it clear that a validation study does not completely validate the test for use with another group of subjects but that it merely increases our confidence



that the test when applied to a group of "similar" subjects will prove similarly valid. Any selection of an existing test to serve a particular purpose (or construction of a new test to serve that purpose) therefore involves validity by hypothesis to a certain extent. The only situation in which we can escape the conclusion that our knowledge of the validity of a test is a hypothesis is the extremely limited one in which the test is validated on the identical subjects for which it is to be used administratively. Since validation of the test involves obtaining criterion measures (which are presumably superior to the test itself and would be used if it were not for the greater time and cost of securing them), the absurdity of using a test which has been prevalidated in this sense becomes immediately apparent. This does not lead, of course, to the absurd conclusion that a test may never be used; rather, it makes clear that when a test is used, its use is based upon a hypothesis in which we have more or less confidence depending upon the amount of research which has preceded its formulation. Our confidence in the test also depends upon the similarity between the research situation and the service-testing situation. Needless to say, this conclusion applies with equal force to all personnel evaluation and prediction devices.

It will be noted that validity by hypothesis departs from the concept of "face validity" in the preceding usages of the term. The first three usages discussed involve a superficial, common-sense similarity between test content and test objective. For example, in validity by assumption the similarity between test and job, without regard to statistical evidence of validity, is taken as sufficient. In validity by hypothesis, the similarity to a test for which there is statistical evidence of validity is tentatively accepted, without regard to its resemblance to the criterion. In validity by hypothesis, no such superficial similarity is assumed. On the basis of extensive previous research, one might legitimately propose that the ability to identify pictured hands as right hands or left hands would be a valid test for the prediction of the ability to read blueprints, although the superficial resemblance between the two tasks is slight. Nevertheless, certain controversies which

have been raised about face validity and the presumed necessity for prevalidating any test before it is used<sup>4</sup> make the discussion of validity by hypothesis appropriate in connection with the other uses of face validity.

Moreover, in validity by assumption, hypothesis, or definition, we are dealing with varying points on a continuum of degrees of certainty. In "assumption" we have, within the scientific frame of reference, no confidence whatever; in "hypothesis" we have varying degrees of confidence depending on the amount, quality and pertinence of the evidence from previous experience; in "definition," our confidence usually is greatest, but—and this must always be remembered—that confidence applies only to the trait or traits actually represented by the test items in relation to the sample and *not* to traits defined in any other way.

### *Summary and Conclusions*

1. This paper has attempted an analysis of the various meanings of the term "face validity." These meanings, although superficially similar, lead to widely different conclusions.
2. The results of the analysis may be summarized as follows. Face validity is variously used to mean that:
  - a) The test bears a common-sense relationship to the measurement objective and therefore no statistical verification is necessary (*assumption*).
  - b) The test sets such a task that the universe of possible tasks (of which the test is a representative sample) is the only practicable criterion and the test is therefore a valid measure of the universe defined in terms of the sample. This implies merely that the test is a valid measure of whatever trait is measured reliably by the test (*definition*).
  - c) In the interests of the acceptability of the test to those most intimately concerned with its use, it is highly desirable that a test possess not only statistical validity,

---

<sup>4</sup> Strangely enough, many of those who insist upon the prevalidation of each written test continue to urge reliance upon other types of selection techniques which numerous research studies have almost unanimously shown to be without predictive value.

but also, as an added attribute, the appearance of practicality (*appearance*).

- d) In the construction or selection of a particular test to be used for a particular objective with a particular group of subjects, recourse is always had to previous knowledge of the effectiveness of the same or similar tests applied to the same or similar subjects for the prediction of the same or similar attributes. On the basis of this previous research, the hypothesis is proposed that this test will be valid for the particular objective. The hypothesis is one which carries varying degrees of confidence: in some cases enough to justify the use of the test immediately, pending further investigation; in other cases so little confidence that such further investigation seems unprofitable. Even after there has been further investigation, however, we are left with a degree of confidence which is somewhat less than certainty, unless we are dealing with the same test, the same population and the same objectives (*hypothesis*).

3. Since the term "face validity" has become overlaid with a high degree of emotional content and since its referents are not only highly ambiguous but lead to widely divergent conclusions, it is recommended that the term be abandoned. Anyone intending to use the term should, instead, describe fully the *concept* which he originally intended to denote by "face validity." Even though writers may not always follow this recommendation, it is hoped that the foregoing analysis will prevent readers from drawing the improper conclusions that have frequently resulted from the indiscriminate uses made of the term in recent years.



## TEST PATTERNS IN THE VOCATIONAL CLINIC<sup>1</sup>

LINDSEY R. HARMON

Veterans Administration

### *I. Some Fundamental Considerations*

THE use of psychological tests in scientific personnel work involves two related yet distinct functions, a "selection" function and a "guidance" function. Several methods of applying tests in personnel selection have been discussed by Toops (10). The essence of the selection problem is to choose the fittest from a surplus of applicants for a given *position*. The "guidance" function begins with the other side of the problem, the *individual* seeking the most suitable job. Tests take their place as quantitative, objective measures of characteristics of importance, or presumed importance, in vocational, educational, or personal adjustment. Other kinds of data, from other sources—records, interviews, personal impressions—round out the picture. An adequate individual survey includes a wide variety of kinds and sources of information about the individual; vocational diagnosis requires a thorough scrutiny of all the available knowledge about the person being counseled.

Some phases of counseling involve essentially an application of the "selection" philosophy, as when the question is, "Can this individual pass the necessary 'hurdles' to qualify for a particular job under consideration?" This fact, however, does not mean that the difference in point of view is nullified. At one time the job, at another time the individual is the focus of attention. "Guidance" and "selection" both seek to bring the two together in a satisfactory adjustment.

These two different ways in which tests can be used in vocational counseling impose different demands on the tests and the

---

<sup>1</sup> The views expressed in this article are those of the author and do not necessarily reflect the official policy of the Veterans Administration

norms used to interpret the tests. In the "guidance" function, where the purpose is to develop a broad, inclusive picture of the individual, the emphasis is necessarily upon the *breadth* of the vocational interpretation that can be made. In the "selection" function, the emphasis is correspondingly upon maximum validity of interpretation for a specific job or job field. For selection purposes a wide variety of norms for various populations is important so that the individual may be compared most precisely with his potential competitors in the various jobs or training programs for which the test is validated. The enormous extent and complexity of the task of providing this kind of data for vocational tests is touched upon by the Staff of the Division of Occupational Analysis, War Manpower Commission (7). These problems will not be dealt with further here except to delineate the difference between them and the problems of interpretation of tests in performing the broader guidance function.

*Norm Groups.*—For guidance purposes, where intra-individual differences are most important, it is apparent that all elements of a test battery must be referable to a common base or the pattern that emerges will be an artifact of the varying norm groups. An example would be a battery in which there are marked sex differences between the tests, some favoring men and some favoring women. If a man taking this battery is compared on all tests with the same "men in general" norm group, the pattern that is found may reasonably be assumed to be a real pattern of the differences within the individual man. Let us assume that this man did outstandingly well on test "X" on which, incidentally, there is a marked sex difference favoring women. If, now, he were to be rated on this test on women's norms, his outstanding capacity of trait "X" would seem to disappear. In analogous fashion, in trait "Y" where he is only average, but the sex difference favors men, he would seem to have an unusual talent if he were to be rated on women's norms. These examples point out the problems of interpretation involved when the several tests of a battery are standardized on varying groups, even though they are all assumed to represent the "general population." The more nearly

comparable are the norm groups, the less is the distortion of the "true" individual pattern. The problem of just what group would be the most useful as a common base or "frame of reference" remains.

There are excellent arguments for having a series of norm groups on whole batteries of tests measuring a variety of mechanical, clerical, and intellectual aptitudes. It would unquestionably be useful to have, as a basis of comparison for a man aged 23, a set of norms on "men in general" aged 23 (or 20-25) and for another man aged 47 a similar "general population" norm for men aged 47 (or 45-50). Similar advantages could be cited for groups varying in their educational level or even in their "general intelligence" level. However, until extensive normative studies have been made to provide a base for such norm tables, the degree of usefulness of such norms must remain largely speculative. Meanwhile, vocational psychologists are limited for the most part to less closely defined "general population" norms. In everyday practice, in a vocational clinic dealing with people of widely varying ages and educational and vocational background, seeking adjustment in occupations of corresponding variability, general population norms, separate for men and for women, serve very effectively if the norm groups for the several tests are in fact comparable. With such a stable base provided, other interpretative problems stand out as more challenging, and as greater sources of error than do problems of normative precision.

In clinical practice, tests do not perform their function alone. They are exploratory tools of considerable precision, but their results are always presented in the context of the life pattern of the individual being counseled. This matter of context is crucial. In the absence of such extensive normative frames of reference as were previously discussed, great importance is attached to the clinical judgment of the vocational psychologist. The context is not, as yet, quantified. Qualitative judgments must be made. Tests are one source of data that must be fitted into the broad pattern of traits and experiences and values and drives that characterize the individual counselee. In many cases the counselor will have formed extensive tenta-

tive judgments before the results of tests are available, by relying on records and interview information. The tests may refute, alter, or confirm the tentative judgments. It is possible in some cases that essentially the same judgments could be made from test data alone. But clinical practice emphasizes rather that scrutiny of a test profile raises questions, the answers to which are sought in other data. Reference to a social history likewise raises questions that tests may help to answer. This interdependence, this knitting together of data from varying sources, is the essence of a sound vocational diagnosis. The life of the individual forms the indispensable context for a maximum validity of test interpretation.

*Tests as Context.*—Sex, age, education, occupational experience and physique, to mention only a few of the more important areas of information, are necessary as proper context for the interpretation of the vocational significance of a given test. Many kinds of information about the individual, however, can best be provided by the results of other tests. The significance of a mechanical aptitude score varies greatly depending on the level of scholastic aptitude of the individual. The practical counseling interpretation of an interest test must be dependent to no small extent upon the knowledge of the aptitudes that will be required to implement the interests shown. Tests thus become part of the context for the interpretation of other tests.

The internal patterning of a test battery, with its revelation of intra-individual differences, provides data of what might be called "context quality." For example, indications of a high degree of mechanical aptitude accompanied by a low degree of mechanical knowledge strongly suggest a lack of interest in mechanics. There might be other reasons than lack of interest, although in the normal course of American life the opportunity for the acquisition of mechanical knowledge is not lacking. The appearance, in such a case, of a low mechanical interest score would acquire added significance and would enhance the validity of the interpretation that "this man's interests are not in a mechanical direction." The one pattern tends to confirm the other; each serves as useful context for the interpretation of the other, just as does life history data.



In similar fashion, the interdependence of personality test profiles and interest test profiles might be demonstrated. And each raises further questions that can be answered satisfactorily only by measures of ability. The dependence of case history data on measures of ability, interests, or personality can be shown in analogous fashion. Test results frequently prove the most useful or even the crucial element in fitting together the "jig-saw puzzle" of life history data—some observed but unexplained bit of deviate behavior may be understood and made meaningful in terms of a personality test pattern. The problem of the kind and amount of test data needed to achieve the maximum of this "context quality" thus comes into focus as a prime element in the individual survey.

## II. *Profiles and Factors*

As closely related to the problem of individual diagnosis as one side of an equation is to the other is the description of the job fields to which the individuals are seeking adjustment. The goal of vocational psychologists and job analysts for a generation has been the description of individuals and jobs in the same set of terms. Repeated attempts have brought clear-cut results in a few areas, but comprehensive descriptions of jobs and workers in common terms have not yet been achieved. One notable success, limited in scope though it is, is the *Physical Demands and Capacities Analysis* of the United States Employment Service. Dealing almost solely with motion demands and working conditions, it is also limited to subjective judgments rather than objective measurement. It has, however, a proved utility, particularly in dealing with the placement of handicapped workers, and in slightly modified form it has become an integral part of the Vocational Advisement procedures of the Veterans Administration.

*Jobs and Traits.*—Measurement of jobs in terms of human traits can only be done by measurement of the human beings who are successful in the performance of those jobs. Examples of this sort of measurement are found in the pioneering work of the Employment Stabilization Research Institute in the development of "occupation ability patterns" (3) and the work

of E. K. Strong (9) in the measurement of interests of occupational groups. These studies provided starting points for further expansion by many workers in the field. The newer statistical techniques, one of which is factor analysis, promise to greatly extend and to refine our knowledge in this area. When factor analyses have been carried to the extent of providing us with job descriptions in terms of primary human traits we shall have made really important strides in the efficient and scientific diagnosis of individual vocational potentialities.

Pending the research which will give the vocational psychologist efficient factorial descriptions of jobs and workers, diagnostic tools from pre-Thurstone research techniques are in use and are not without a considerable practical efficiency. Broad groupings of jobs into functional categories such as "professional," "clerical," "skilled," "semi-skilled," "sales," etc., correspond roughly with broad groupings of traits well recognized by vocational psychologists and are not in conflict with such factor studies as have been made. Years ago Thorndike spoke in terms of abstract, social, and mechanical intelligence. Experimental studies since that time have isolated as relatively unique traits such abilities as academic, clerical, mechanical, and manual aptitudes. These groupings have clear functional significance, perhaps most clearly described in the literature of vocational psychology by Bingham (2). Occupational ability profiles in terms of these traits have been developed and, few though they are at present, they provide invaluable data for the vocational counselor. They serve first in the broader "guidance" function, in indicating the general direction and level of the individual's vocational potentialities, and in indicating fields in which the individual's abilities may best be utilized. In other instances they will have a more specific value in the later phases of counseling, in performing the "selection" function of tests to help check the individual's capacity for a specific job.

In these two phases or functions of counseling—the broad general orientation and the closer, more precise checking—it is important that the initial survey be comprehensive enough to

omit no significant area of adjustment potential. Unless this is done, there can be no assurance that the next or final approximation and decision as to a job goal is within the field of the individual's greatest potentiality. Intellectual, clerical, and mechanical aptitudes; vocational interests; and personality measurements—these constitute a minimum basis for an adequate psychological survey. There is no supposition here that personality functions independently of interests nor that interests are unrelated to aptitudes. The separate consideration which they are to be given is rather a matter of convenience in the organization of the data. Three sets of profiles are considered as prime requisites: aptitudes, interests, and personality. The integration of the data to achieve the maximum of "context quality" will then be considered.

*Scholastic Aptitude.*—If aptitudes in the areas of academic, clerical, mechanical and manual abilities were simple unitary traits, four tests are all that would be needed for a minimum ability profile. But we know that these traits are complex, and to have measurements adequate enough for individual diagnosis it is frequently necessary to have two or more tests in each of these four areas. Examples in the area of academic ability or "general intelligence" are the simple profiles or patterns provided by the Q and L scores of the *American Council on Education Psychological Examination* or the Verbal and Performance Scales of the *Wechsler-Bellevue Intelligence Test*. And in each case the test may be, and frequently is, broken down further into its component sub-tests for further diagnostic detail. Details of the interpretation of these particular tests are beyond the scope of this paper. The important point is that the vocational psychologist usually seeks through a study of the pattern of such tests to throw light upon the difference between a verbal-abstract-academic kind of ability of the individual and a quantitative-mechanical-concrete kind of ability. More thorough exploration of these trait-patterings is then sought through the use of specific aptitude tests.

*Clerical Aptitude.*—Most closely related to academic ability, both from the standpoint of test intercorrelations and practical vocational requirements, is clerical aptitude. Insofar as a

special ability is involved, the requirements seem to be rapidity and accuracy in the handling of details, particularly of figures and names. In any but routine clerical jobs, of course, intellectual capacity is usually determinative of success and promotability. The special aptitude for detail remains important, however, even at the higher levels of clerical operations. One of the well-known and satisfactory means of measuring these traits is the *Minnesota Vocational Test for Clerical Workers*. The difference between the two subtests, Number-Checking and Name-Checking, as well as the absolute magnitude of the scores, is frequently worthy of note as a "straw in the wind" of individual diagnosis. The original research upon which the test was based (1) as well as subsequent research by the United States Employment Service (8) indicate differential patterns of ability among various groups of workers. The differences are chiefly along such lines as the level or routineness of the work, or whether it deals primarily with computations or with verbal material. Study of these differential norms will provide the counselor with invaluable clues in extending and enriching his interpretation of the pattern of the subtests as well as of each subtest separately.

*Mechanical Aptitude.*—Measurements of mechanical aptitudes have probably been more thoroughly developed than have those of any other group of special aptitudes and several good tests that are adapted to the adult level are generally available. The outstanding component of basic mechanical aptitude appears to be what the factor analysts have described as a "spatial factor"—the capacity to perceive, remember, and manipulate mentally, sizes, shapes, and angles. Another important type of measurement is that involving in addition to space perception, mechanical knowledge. Examples of the former type of test are the *Minnesota Spatial Relations Test* and the *Revised Paper Form-Board Test*. A good example of the latter type is the *Bennett Mechanical Comprehension Test*. Patterns of scores from these three tests are frequently of diagnostic value. For example, the Paper Form-Board is more abstract or academic, the Spatial Relations test more concrete and "practical." These differences repeatedly appear in tables

of intercorrelations and are apparent also in clinical practice, where bright people tend typically to excel on the paper test and those with less academic ability and background do better on the performance test (Spatial Relations). The difference between both of these tests and the Bennett, which is more profoundly influenced by mechanical experience and training, as shown by the validation data, is also of value. Time and again in counseling experience, score differences appear which correspond with either measured or expressed mechanical interests or experience with mechanical things. This pattern may frequently be useful as an aid in diagnosing the significance of the difference between expressed and measured interests, as was mentioned previously in Part I of this paper. This use of test patterning closely resembles the "Objective" methods of measuring interests in terms of acquired knowledge described by Fryer (4) or the use of the *Michigan Vocabulary Profile*.

A word of caution is in order at this point. The reliability of a difference between tests is necessarily much lower than the reliability of either of the tests taken separately. This is true regardless of the tests being compared. Two decades ago Kelley (6) called attention to this fact and warned of the dangers of using ratios (Accomplishment quotient) of test scores. A straightforward study of profile differences, although subjective, avoids the false importance that is attached to the "objective" calculated ratio. The statistics of pattern analysis, as Toops (10) has pointed out, are techniques as yet undeveloped. But the observation of test patterns, considered in relation to other data, adds a clinical detail always worthy of attention.

With experience in studying these patterns the vocational counselor comes to develop a "feel" for consistency of test profiles, for patterns that fall in line with normal expectations. A constant study of tables of test intercorrelations is essential in the development of this aspect of clinical judgment. Profiles that do not conform with expectations raise questions that the careful counselor will wish to answer, if possible, before proceeding with advisement. Simplest among these questions is that of scoring accuracy. In spite of the precautions taken

to insure accuracy, errors will occur, and what appears to the counselor as a "discordant note" in a test profile will not infrequently be traceable to an error in scoring or norming procedures. Sometimes test-taking attitudes are revealed—attitudes that may affect a score on one test more than another, such as an aversion to paper-and-pencil tests but a liking for performance tests. Sometimes there is no ready explanation of the pattern. In such instances, it is necessary to regard the test results with greater tentativeness than usual, while seeking by means of interviews or further testing to find out what may lie back of the observed departure from the "norm."

*Manual Dexterity.*—Manual abilities, as many studies have shown, are highly specific. That is, the correlations between various manual or psychomotor tests tend to be low. In the absence of specific validity studies, therefore, the vocational significance of tests of manual dexterity is rather tenuous because the same individual may score very high on some manual tests and very low on others. Experimental studies, such as those of the Employment Stabilization Research Institute (3) have indicated also that for many skilled occupations manual dexterity is definitely of tertiary importance. The mental, rather than the manual processes involved in skilled work, are of determining importance. This is a point difficult for an untrained person to grasp. It is easy to "see" the importance of deft finger motions in observing a skilled worker at his task. Ergo, manual dexterity is involved and manual dexterity tests should be good predictors of ability. The fallacy lies in the fact that the observation is necessarily superficial—it is impossible to see the mental processes that guide the fingers and to observe directly that it is the mental or mechanical skill, and not the speed of finger motion, that is determinative of proficiency. But most manual dexterity tests measure only routine speed, the rapidity of repetitive operations. In jobs where repetitive speed is important, such as semi-skilled assembly operations, the importance of dexterity is easier to demonstrate. The high degree of specificity of manual abilities, however, requires that an interpretation of the results of manual tests be closely guided by specific experimental evidence, which severely limits the generalizations that may validly be made from any

manual test or series of manual tests. Future research may widen the horizon of pattern-interpretation in this area; meanwhile the careful counselor will be cautious in his predictions.

*Vocational Interests.*—Interest tests fall into two broad groups: those in which items are related directly to jobs during the original test-construction procedure, and those in which a few relatively homogeneous scales are provided, which must subsequently be validated by experimental studies. Outstanding in the field, the *Strong Vocational Interest Blank* is an example of the former procedure. Its scales relate directly to specific jobs, and give scores in terms of occupational norms. The second type of test is well exemplified by the *Kuder Preference Record*, whose nine scales, derived by statistical techniques, are not directly related to specific jobs. The interpretation of the pattern of these scales is dependent upon subsequent research which provides "occupational profiles." A few such profiles appeared with the original test. Constant research since then has accumulated data for many more, which have recently been prepared and published. Unless such aids are provided, interpretation is dependent upon arm-chair judgment rather than verifiable fact. The same caution is even more true of tests constructed by "expert judgment" of the significance of items, and the history of vocational psychology is strewn with the demonstrated failures of such tests.

In counseling an individual it is relatively seldom that one may utilize in a one-to-one manner the results of vocational interest tests, as only a few of the many occupations with which the counselor and his counselees are concerned will be specifically covered by the test-validation procedure. It is necessary therefore to consider the test results in terms of job families that may be indicated by the test patterns. Factor analysis of the 36 occupational scores yielded by the Strong Test has provided some experimental data about which job families, in terms of interests, may be organized. As a result, counselors speak with more confidence of interests in "business detail occupations," "scientific-technical occupations," "social service occupations," etc., rather than in specific jobs, in working out the first general orientation or direction of the individual's interests or motivations. When data bearing on the

specific job choice are available, it is highly useful in the later processes of applying the "selection" philosophy to check the suitability of the chosen job. Pending the development of large numbers of occupational keys or occupational profiles, the counselor will have to continue to depend upon vocational interest "families" of occupations and sound clinical judgment based upon a careful study of the extensive literature of vocational interest investigations.

*Personality.*—A diagnosis of personality is at once the most fascinating and most difficult problem of the vocational clinician. From the pioneering study of Woodworth in World War I, which produced the *Psychoneurotic Inventory* until the present there have been literally scores of attempts made to measure personality traits of importance in vocational adjustment. Some scales, such as the *Heidbreder Introversion-Extraversion Test*, have been based on a clearly defined theory of personality. Many, if not most, have been found to be worthless because the traits they measure cannot be shown to have any vocational or clinical significance. Only a few scales in the personality field have been based upon empirical evidence of the capacity of the items to differentiate clearly defined occupational or clinical groups. Research in this field is peculiarly difficult, expensive, and time-consuming, and this fact above all others has probably been responsible for the dearth of usable scales. In spite of the difficulties, however, some useful results have emerged, and vocational counselors are no longer dependent solely on their subjective judgment of personality traits.

Among the better-known instruments developed by group-differentiation techniques is the *Bell Adjustment Inventory*. Without attempting any abstruse theoretical formulations, this test seeks to measure directly the adequacy of over-all adjustment in the fields of home, health, social, emotional, and occupational relationships. It is simple and easy to use, and subject to a minimum of misinterpretation. Its chief limitation, in clinical practice, is that its scales have little "subtlety," and the results obtained are only such as the individual chooses to reveal about himself.

More complex in interpretation, the *Minnesota Multiphasic Personality Inventory* has already demonstrated a high degree



of practical usefulness in the field of vocational advisement (5). Its scales, based on a differentiation of clinical groups of hospitalized psychiatric patients, are chiefly useful in dealing with the more severe forms of maladjustment. An outstanding feature is the fact that the test is capable of revealing traits of personality of which the individual himself is unaware, or even may be attempting to conceal. Further research now in progress gives promise of increasing its usefulness in this area.

### *Summary*

The discussion of test patterns that has been undertaken here is intended only as an exposition of fundamental principles illustrated by a few basic examples. For a more comprehensive treatment of the interpretation of test patterns, numerous examples with case histories would be required. The articles and books cited provide some of these examples and a quantity of experimental data and scientific investigation upon which valid interpretations of test patterns may be based. The counselor or vocational psychologist who seeks to draw maximum value from the tests he is using will find that the study of the experimental literature, combined with an application of psychological principles to the diagnostic problems he is facing in his daily work, will be greatly rewarding. The threshold has barely been reached, and certainly not yet crossed, into the quantitative study of test patterns. Until statistical tools for that work have been provided, the counselor will be dependent primarily upon the degree of qualitative clinical insight and judgment that he may acquire.

Tests used as an aid in vocational diagnosis serve the function of rounding out a comprehensive picture of the individual counselee. Their data are factual, objective, and relatively precise as compared with most social history information. The utility of test data are, however, largely dependent upon the adequacy and comprehensiveness of the context in which tests are used. Test information itself, properly organized and evaluated, becomes useful context for the interpretation of other tests as well as for the interpretation of case history and behavior data.

A test profile provides a graphic pattern of traits—aptitudes, interests, and personality—which forms a vitally important aid in vocational diagnosis. The physical arrangement of the profile translates the pattern of traits within the individual into a visual pattern that helps the counselor to integrate the results of a wide variety of measures into one meaningful whole. It enables the counselor to make a series of measures more useful than the several tests taken individually. It diagrams assets and liabilities and serves to facilitate a general orientation to types and levels of jobs. More closely examined, it provides by its patterning numerous “clinical clues” which become important in a more searching diagnosis. It does not supplant the history or interview, but rather lends to both an added significance, and is itself dependent upon them for its fullest interpretative value.

#### REFERENCES

1. Andrew, Dorothy M. and Paterson, Donald G. *Measured Characteristics of Clerical Workers*. Minneapolis: University of Minnesota Press, 1934.
2. Bingham, Walter Van Dyke. *Aptitudes and Aptitude Testing*. New York: Harper and Brothers, 1937.
3. Dvorak, Beatrice Jeanne. *Differential Occupational Ability Patterns*. Minneapolis: University of Minnesota Press, 1935.
4. Fryer, Douglas. *The Measurement of Interests*. New York: Henry Holt and Company, 1931.
5. Harmon, Lindsey R. and Wiener, Daniel N. “Use of Minnesota Multiphasic Personality Inventory in Vocational Advisement.” *Journal of Applied Psychology*, XXIX (1945), 132-141.
6. Kelley, Truman Lee. *Interpretation of Educational Measurements*. Yonkers-on-the-Hudson: World Book Company, 1927.
7. Staff, Division of Occupational Analysis, War Manpower Commission. “Factor Analysis of Occupational Aptitude Tests.” *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, V (1947), 147-156.
8. Stead, Shartle and Associates. *Occupational Counseling Techniques*. New York: American Book Company, 1940.
9. Strong, E. K., Jr. *Vocational Interests of Men and Women*. Stanford University: Stanford University Press, 1943.
10. Toops, Herbert A. “Philosophy and Practice of Personnel Selection.” *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, V (1945), 95-124.

# THE ERGIC THEORY OF ATTITUDE AND SENTIMENT MEASUREMENT

RAYMOND B CATTELL

University of Illinois

## 1. *Dynamic Intensity and Dynamic Direction*

THERE is a well-known axiom of science which was expressed by Lord Kelvin in the words: "When you can measure what you are speaking about and express it in numbers, you know something about it." The attempts of dynamic psychology to attain this high plateau of scientific progress have so far largely failed. They have fallen on the one hand into formalities of measurement, often spurious, and generally with trivial dynamic situations, and on the other, into verbal, generally clinical, conceptual elaborations too lacking in precision of thought to be subjected to experimental verification. Our purpose here is to propose a remedy in the form of certain clear theoretical conceptions and a practical method.

Our point of departure into psychodynamics will be the measurement of attitudes, which is already the center of much discussion (2, 4, 6, 13, 18, 20) having been simultaneously traversed by sociologists and by students of normal and abnormal personality. The sociologists, however, who have been responsible for practical steps, have adopted a practice inconsistent with dynamic psychology, and the present writer has insisted for some years (5) that the meaningful measurement of an attitude requires more indices than current practice includes. For though social problems may be handled with momentary success by these jejune concepts, social predictions must be linked in the end with the dynamics of the individual if any realistic science is to emerge.

The first and basic conception in the theory now proposed is that *an attitude is a vector, definable by direction as well*

*as magnitude, and further by point of application (object) and stimulus situation.*

Current attitude measurement theory assigns an object and a single quantity—strength—along a pro-con axis, to an attitude. Attitude scales are, therefore, at present devised on the theory that one must be either “for” the church, the Germans, the Jews, birth control, etc., or “against” them, neglecting all the other points of the emotional compass through which attitudes can range. Recent developments of attitude measurement methods, notably those by Guttman (13, 14) with Stouffer and by Cantril (4), it is true have begun to use a second dimension, of intensity or certainty, but their theory is quite distinct from that proposed in the present paper, as will be illustrated later (p. 241).

Although attitudes can attach to either concrete or conceptual objects, the above argument can be directly illustrated by a concrete object—the pretty Miss Doe. Attitudes to this object may spring first from the innate propensities or *ergs*,<sup>1</sup> taking the form of anger, fear, sexual attraction, parental protectiveness, gregariousness, or even, in the case of a cannibal, of hunger! Secondly, they may spring from the innumerable secondary, derived emotions expressing the individual's metanergic structure as it derives from the ergic combination patterns forced by social institutions. Surely, it would defy the attitude test designer to say whether the man attracted by Miss Doe's physical beauty, or the cannibal viewing her as a culinary object, are “for” or “against” her!

## 2. *The Definition of the Total Attitude*

Our thesis, therefore, is that the immediate need of attitude measurement theory is to find means of defining attitude direction—a task side-stepped by those who have concentrated on

---

<sup>1</sup> An *erg* is “an innate disposition . . . to acquire reactivity to some situations more readily than others, to experience a specific emotion therefrom and to enter on activity which ceases more completely at the attainment of one specific goal than at other situations. Associated with it is an innately preferred behavior in reaching the innately preferred goal” (7, p. 195). The term *drive* means so many things to so many people that I shall throughout this article stick to the operationally defined term *erg* (6, 7), despite its being familiar at present to only a small group. This is desirable also because, as the result of our discussion shows, the pursuit of exact attitude measurement becomes a problem in the delineation of ergic structure.

measuring strength only.<sup>2</sup> Now direction theory involves both finding and fixing. That is to say, as with any natural phenomena, we have to find what exists in nature, i.e., what coordinates are needed, and then we have to work out a labelling and measuring scheme to fix any given attitude.

It is implicit in the whole usage of dynamic psychology that what we mean by direction is an intention or goal of action. The proper beginning of attitude study, therefore, is in a search to discover ways of formulating dynamic directions. We shall begin this search with a formal statement of what we consider is comprised in an attitude, sentiment, or ergic trait. This formula, which includes the basic emphasis on direction, states that the following is the minimum definition of an attitude or ergic trait:

"In these circumstances I want so much to do this with that."

Attitude theory has concentrated exclusively on (1) "so much,"<sup>3</sup> paying little attention to (2) "I" (3) "that" and ignoring or assuming completely (4) "In these circumstances" and (5) "to do this."

Dynamic direction, it has just been said, needs to be defined by some set of coordinates naturally inhering in psychodynamic data. Theoretically, there seem to be some three ways in which coordinates of intention could be obtained, as follows:

<sup>2</sup> Fortunately the methods of measuring strength, which have been worked out with great explicitness and practical success by Thurstone (20, 21) and others (12, 13, 16, 18), are then still suitable to measure the strength of the direction-defined attitude.

<sup>3</sup> In parenthesis it is necessary to say that the present article will not be much concerned with the actual experimental form of the measurement of strength, i.e., whether it be by direct or indirect verbal opinion scales, as used by Thurstone or Guttman, or such physiological devices as the P.G.R., or objective behavior tests. These can readily be sought, shaped, and improved when the true framework for validation has been reasoned out and verified. Whatever the measurement form, it has to be validated in every case ultimately against the amount of *behavioral activity toward the goal stated in the attitude*. That is to say, an attitude is potential behavior. When we measure an attitude we are measuring a dynamic trait in its suspended state. An attitude is, therefore, a *mental structure*, like a sentiment, erg or other dynamic trait, which functions from time to time in a certain pattern of behavior. It may sometimes be convenient to refer to an attitude by the conscious concomitants of its functioning—the primary or secondary emotions, as when we speak of an attitude of fear or hunger above—but, naturally, this is a loose alternative description since only the associated behavior can be measured. In short, the magnitude and direction of an attitude is ultimately measured by the magnitude and direction of the performance when the attitude is allowed to function.

1. By arbitrarily selecting certain practical goals of behavior, usually in connection with a social institution, e.g., the good of society, the economic standing of the individual, the individual's mental health, the growth of the church, the destruction of certain objects, the increase of this or that, it would theoretically be possible to discover how far any given line of behavior contributes to each of these, i.e., to give its projections on these coordinates.

2. By factor analysis or other methods applied to the social and physical environment one might discover the nature and number of the independent directions of social change. One would next need to discover how far any particular (intention of) behavior contributes to each of these directions of change. This can be called the "social emergents or social consequences basis" for attitude evaluation.

3. By factor analyzing a very varied set of individual attitudes, by the ordinary R, Q, or P techniques (6), one might arrive at a limited number of dimensions of dynamic purposes of the individual. This we can call the "organismic, dynamic" basis or "dynamic composition" basis.

The first two must be rejected.\* The third is practicable

---

\* The first fails (1) because of the unknown relationships among the arbitrary coordinates, e.g., favorableness to morality and favorableness to economic advance may or may not be independent coordinates; (2) because we do not know whether the arbitrary coordinates exhaust the dimensionality of the true space; and (3) unless the projection of a given attitude on these coordinates is to be worked out on a dynamic basis (in which case one would do better to go over entirely to the ergic basis here described) one is involved in the far-fetched search for a sociological connection, as in system 2.

System 2, though of great theoretical interest, fails because it attempts prematurely and in an arbitrary fashion to link two fields, those of individual dynamics and social dynamics. If the analysis of individual dynamics of attitudes were followed out as here advocated there would still remain (a) the enormous task of factorizing the dimensions of social change and (b) the equally enormous task of finding how individual dynamic attitudes affect social change. The first of these has only recently been methodologically sketched (7) and the second is likely to develop only as a very complex set of corollaries after the dynamic analysis here suggested has reached a certain experimental maturity.

Current pro-con attitude methods, even when they constitute a true attitude direction by being pro-con a course of action instead of pro-con an object, are vitiated by a confusion of thought between methods (1) and (2). Sometimes they set up arbitrary coordinates as in (1), e.g., "against conscription," "against republicanism"; but when as in (2) they speak of independent dimensions of social change organic to social structure, there is neither proof that they are independent, nor proof that the individual's attitude in fact contributes to the direction of change accepted as the coordinate. For example, it is not certain that an attitude against armament expenditure is in fact pacifistic, or that an intention to have church X supported by state funds would in fact be favorable to the growth of the church.

but has not previously been proposed or tried; for the existing, published factor analyses are on collections of scales pre-selected to have a pro-con direction. If these scales dealt each with a particular course of action, the correlations would show that approval-disapproval has different dynamic quality direction in different situations. But they are almost invariably pro-con objects (institutions) and, therefore, merely yield groups of *objects* liked or disliked instead of groupings of the possible dynamic trends in attitudes generally. Let us now, therefore, examine systematically how the third method would work if properly applied to the end of discovering dynamic factors or dimensions in individual attitudes.

### 3. *Subsidiation and the Dynamic Lattice*

Before the probable result of a factor analysis of dynamic traits can be envisaged accurately enough to produce an effectively designed experiment, it is necessary to have in mind a realistic picture of the total dynamic life of the individual. We shall attempt to do this by means of the concept of a *dynamic lattice*. This aims to bring out the relations between underlying dynamic structure and overt behavior manifestations. The example of a dynamic lattice presented in Figure I is intended to serve both as a generalized statement of metanergic structure and as a sketch for the associated discussion of concrete examples.

The dynamic lattice seeks to represent a particular or a typical person's dynamic relations to the interest objects of his everyday life. The symbolism convention is that each channel represents a line of action, an attitude toward one of these objects. The object may be a concrete object, a concept, or an activity. One object can obviously be the recipient of several attitudes or intentions, each appropriate to one of a limited number of situational circumstances. An attitude has to be defined by direction, strength, object, and circumstance (stimulus). The object is represented by a circle, what the person wants to do with an object, i.e., the attitude direction is shown by the ultimate destination of the channel, while strength can be shown by its thickness (but this diagram

has not been complicated by any precise attempt to render strength). Circumstance (stimulus situation) is implicit in the pattern of the lines among the objects. Since we are dealing with relatively fixed personality traits and attitudes, it is assumed that the life circumstances on which they depend are relatively fixed. Nevertheless, it would be possible to indicate specific circumstances by symbols (as 1, 2, 3, etc.) at the outset of the attitude line, always providing that we remember that the larger part of the stimulus situation is usually the object itself upon which the individual operates.

Essentially the diagram brings out the typical, but frequently overlooked, interflux of subsidiation chains of attitudes. (It will be found convenient in this discussion to use Murray's (17) term "subsidiation" for the generally recognized dynamic sequentialness of attitudes and the term interflux for the convergence and divergence of purposes with respect to any single course of action.) Although an attitude is completely defined by the formulation above, one must not forget that the "direction" part, i.e., "to do this with," includes the later chains in the subsidiation sequence. For one does the immediate action always "in order that" one may be able to do something further.<sup>5</sup> What one is "doing" at the given moment, therefore, connotes the ultimate object of the action, i.e., when a man picks up a knife and fork he is "eating." The physical description of the behavior is enough to fix the attitude direction, but the final intelligible definition by dynamic coordinates awaits knowledge of the further subsidiation. The formula of an attitude thus has to include its setting, and "to do this" must be understood to include "in order to." This chain of subsidized attitudes ends, of course, when no farther goal can be discovered, i.e., when the activity or object, e.g., eating (except for a nutritionist!) is a biological goal sufficient to itself. These are, by operational definition, *ergic* goals.

Attitude and sentiment will now be used in the more precise sense introduced in a previous paper (6), according to

<sup>5</sup> Sometimes the further action centers wholly or partly on the same attitude object, as when a young man says, "I will behave charmingly to this girl so that she will let me dance with her."



which an attitude<sup>6</sup> is a purpose subserving and expressing a sentiment. For example, because of his patriotic sentiment an individual may take up a certain attitude about the use of the national flag in advertisements. In other words, this means that attitude objects come early in the subsidiation sequence, sentiments later and ergic goals last. Other accompanying differences are listed below,<sup>7a</sup> but they are all of degree; a line cannot be sharply drawn. (The nomenclature situation is best summarized by saying that attitude, sentiment, and erg correspond to twig, bough, and trunk in describing the "subsidiation" of a tree )

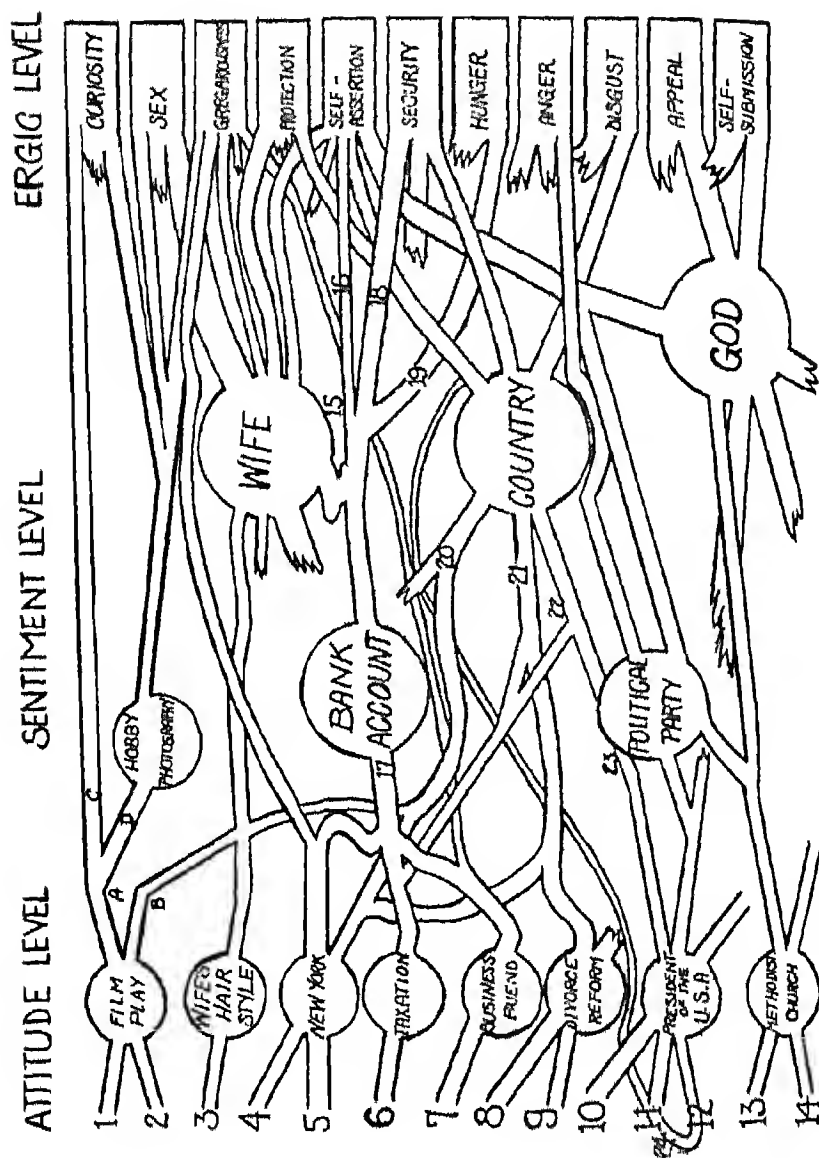
A second differentiation, within both attitudes and sentiments, is important. McDougall (17) defined<sup>7b</sup> a sentiment as a compound of dynamic purposes centering on one object, as on "country" in Figure 1. This is, at least at first sight, not a simply operationally definable reaction, as is any one of the "component sentiments" 20, 21, and 22, in Figure 1. Each object in its time plays many parts. What we want to do with a given object varies according to a limited number of situations—the few situations that actually stimulate us among the limited number in which we actually encounter the object in our lives. The attitudes that we can actually measure, because of their singleness of intention, are these *component attitudes* or *component sentiments*. The question of what meaning, use or reality can be given to a measure of the indubitably real abstraction possible from the complex of attitudes centering on the object (and which, by contrast, we will call the *total sentiment*) must be left until later in this article.

The realities of dynamic structure with which our method has to cope need now to be brought freshly to mind by some concrete illustrations in Figure I—of attitudes to a film play, to

<sup>6</sup> There is also a generic use of attitude to indicate *any* dynamic trait, but this is scarcely to be encouraged in view of the availability of the latter term

<sup>7a</sup> Sentiments have (1) stronger associated emotions, (2) less distance to the ergic goal, (3) greater total dynamic investment, (4) lesser complexity of ergic constitution, (5) more permanence, because of objective conditions (see below), (6) lesser intellectual complexity

<sup>7b</sup> I take it that "sentiment" will not be misunderstood in any loose popular sense of "idealistic inclination." A reader under such misconception has said "in determining attitudes position in life is more important than sentiments" Obviously sentiments and attitudes are structures arising (partly) from the individual's social position, but position is not structure



higher taxation, to a business associate, etc. The man's attitude —17— to his bank account has the direction that *he wants to increase it*. The lines of subsidiation to the right indicate that he wants to do so *in order* to protect his wife —15—, to satisfy self-assertion —16—, to assuage his fear of insecurity —18—, and to satisfy hunger —19—. The lines of subsidiation to the left indicate that this attitude or sentiment to his bank account is served by an attitude of annoyance toward higher taxation —6—, by an intention to keep company with his business friend —7—, and by an attitude of avoidance to New York, where he spends too much money —5—.

Multiple attitudes to an object are illustrated by the film play, which he would like to see himself (attitude in circumstance '2'), from curiosity and because of the help it would

Figure I. The Dynamic Lattice

*Attitude*

1. "I think this movie should not be shown to the public."
2. "I think it would be good for *me* to see this movie personally."
3. "I would like my wife to keep this attractive hair style."
4. "I think more money should be spent improving the defenses of New York."
5. "I do not want to visit New York more than is necessary."
6. "I want taxation to be reduced."
7. "I want to see a lot of my business friend X."
8. "I want divorce reform to be freely debated in the press."
9. "I am opposed to further legislation favoring divorce reform."
10. "I think the President should not waste his time talking to all and sundry."
11. "I wish the President would pay more attention to the party machine."
12. "I want to vote for a democratic President."
13. "I wish the Methodist Church would adopt certain theological dogma."
14. "I intend to spend more time in Methodist Church activities."

A straight channel through the object represents a single attitude to the object. The sharp angular change at the end of such a channel corresponds to the "in order that" which precedes the next attitude and course of action, as at *A* and *B*. Most curves and gradual direction changes in the diagram are, however, nothing but necessary adjustments to a two-dimensional representation.

give to his photographic hobby (itself a sublimation of sex, gregariousness, and curiosity), but which he does not think should be generally shown (attitude in response to '1') because it would be bad for the morale of the country. Here are two attitudes which do not conflict because they spring from different total situations; but it is a constant and inherent feature of attitude formation that actually conflicting attitudes to an object will arise, from the demands of different sentiments and ergs. We are dealing here, however, only with the final structure of adjustment, when such conflicts have been substantially resolved. For example, the attitude to taxation, insofar as it derives from the patriotic sentiment, might be that it should be increased, whereas his personal sentiment to his bank account may require the attitude that taxation be reduced. These will resolve by mutual inhibition into some attitude of middling strength, unless, by intelligent insight or lucky accident some different direction of attitude<sup>8</sup> is found which gives expression to both needs simultaneously.

The problem of faithfully describing the real complexities of motivation in daily life by the symbolism of a dynamic lattice, and of reducing it to essential concepts and principles, is too wide to be completely handled in the present article. Because an appreciation of the whole is relevant, however, to a proper understanding of what is being measured in the individual attitude, we shall make the following preliminary generalizations:

1. The general direction of subsidiation is from a large number of attitudes (lesser, ancillary purposes) to a smaller number of ergic goals.
2. Dynamic subsidiation paths form a lattice structure, both converging and diverging with respect to objects (and to courses of action).<sup>9</sup>

<sup>8</sup> This is, perhaps, an apt point to emphasize that different behavioral ways of obtaining the same end do not constitute differences of direction. Any courses of action which satisfy the same combination of purposes have the same direction. Consequently, in this example a new *direction* is taken only if a course of action with regard to taxation is found which causes changes in the satisfactions.

<sup>9</sup> This pattern is simply and solely a result of the nature of the external world, which permits (a) the same object to serve different needs in different situations, (b) the same course of action in respect to an object to serve simultaneously different ends, and (c) the same course of action in respect to an object to be served by several distinct prior ancillary attitudes.

3. The general direction of subsidiation is broken by many cross currents (interflux), so that the assignment of ergic goal distance (drop in goal gradient) to a given attitude course or object is not simple. For example, attitude 12 serves 23 ([in order] "that I may have my political party predominant"), which subsidates to component sentiment 22 ([in order] "that my country may be preserved as a democracy"). But 22 now subsidates party to 24 ([in order] "that the ballot may be preserved") which is a necessary prerequisite attitude to the opening attitude of our series, number 12. Such retro-active attitudes as 24 may be comparatively rare, but not as rare as our diagram, for the preservation of clarity, has had to make them.<sup>10</sup> Another typical break of the orderly sequence is seen where an attitude to a certain object in part satisfies ergic needs directly (as at *C*) and in part leads to satisfaction only through a long sequence of attitudes and sentiments (as to some extent at *D*).

4. The general direction of subsidiation is opposite to the temporal order (and difficulty) of learning the attitudes in the chain. The attitudes and finer discriminations acquired last and with most difficulty (through blocking of simpler satisfaction) occur earliest in the subsidiation sequence, i.e., have the greatest ergic goal distance or goal gradient.

5. Sentiments will tend to be more constant landmarks of personality than attitudes.<sup>11</sup>

<sup>10</sup> These principles could be illustrated more simply by animal experimentation, as follows: (1) An animal may be trained to lift three latches, in any order, in order to open a door to get food. His "attitude" to each of the latches is that a certain thing must be done with them. His "sentiment" to the door is that it must be opened. (2) Convergence, in the transfusion of dynamic trends, rarely occurs in existing experiments because a simple learning motive—food or escape—is generally employed instead of normal multiple motivation. However, if a maze section were learned in connection sometimes with hunger and sometimes with escape motivation the attitude "to turn right at this point" would constitute such an "overdetermination" or "convergence." (3) The animal takes the shortest possible route. Wider detours, or pressing lever *B* before pushing bar *A*, result only from the failure of the simpler responses and are learned after the simpler response has first led to some temporary success, i.e., late learning is early subsidiation position.

<sup>11</sup> Because of (a) the merely circumstantial fact that the objects of sentiments tend to be more permanent than the associated conditions of their environment, e.g., the sentiment to one's son remains little changed, the attitudes by which the sentiment is served change with his school, his age, one's financial condition, etc. This greater permanence is not entirely environmentally determined. The greater investment in sentiments in turn causes one to keep them more constant. Whatever degree of functional autonomy (2) or disposition rigidity is attained with age will tend to invest

Since the objects<sup>12</sup> of attitudes and sentiments are fulcrums of adjustment, the measurement of attitudes for direction and intensity in the conceptual framework offered here should provide one of the most profitable ways of studying the dynamics of conflict resolution. Conflicts arise because (a) the satisfactions of one component attitude to an object may conflict with that of another, through demanding an incompatible use of the object; (b) the satisfaction of an attitude to one object may conflict with the satisfaction of an attitude to another object. Presumably the organism proceeds by trial and error to the greatest possible total satisfaction. Comparison of the strength of particular aggregates of attitudes before and after certain re-adjustments and re-distributions have taken place would seem to offer a basis for a precise index of the relative goodness of adjustment of individuals or groups under different conditions.

6. The above dynamic lattice deals with subsidiation courses as they actually exist, which in an appreciable percentage of dynamic traits will not be the same as those apparent to the individual's own consciousness. Repression, rationalization, "logic-tight compartments," and other defense mechanisms blot out links of subsidiation and render the method of asking "why this attitude?" impracticable.

#### 4. *The Possibility of Ergic Coordinates from Factor Analysis*

From the above dynamic lattice, which merely systematizes for the purposes of more exact experiment what are the com-

---

the sentiments because they are of longer standing and possibly also because of their greater emotional investment and nearness to ergs. (c) The subject may himself make cognitive reconstruction of attitudes without any major upheaval of personality, e.g., he may decide that his sentiment of love for his home and desire to beautify it is better served by reducing his insurance and painting the house, i.e., his attitude to insurance is changed.

<sup>12</sup> In employing the dynamic lattice the questions should be raised as to whether "objects," rather than "courses of action" or "stimulus situations" are the more convenient points of reference and analysis. If one and the same object changes its meaning according to the dynamic context and is, as it were, a mere pawn in several different attitude courses, might it not be better to choose one of these other aspects by which to pick out and name the dynamic unit? These issues are considered three paragraphs below in connection with the meaning of sentiment. Meanwhile it is perhaps sufficient argument for picking out attitudes principally by object "labels" that objects are the common referents for many people and the items whose movement relative to the individual we most desire to predict.

monly accepted observations of dynamic psychology, we may attempt to infer the general nature of the coordinates of dynamic direction and the means by which they may be extracted. If ultimately the multitudinous common attitudes end in quite a limited number of drives or independent ergic goals, these should provide the reference system required for defining the direction of any attitude.

For immediate practical purposes some psychologists may be tempted to adopt the independent coordinates suggested by Freud's subtle but fragile analysis of instinctual forces. Others may prefer McDougall's shrewdly drawn list of biological propensities or Murray's more elaborate and speculative nomenclature of needs. The degree of reliability of these approaches is indicated by the amount of agreement among them. They involved remarkable insights, and pursued the issues as far as the human intellect could without the invention of new methods.<sup>13</sup> But, except where wearied verbal contention has stagnated in certain vague outlines of "accepted truth," there is little of the finer analyses of these pioneers that cannot be trampled into oblivion by hordes of less intelligent and less fair-minded followers. The findings have the pragmatic truth of some degree of practical usefulness, but they do not bring to dynamic psychology the accuracy of prediction or the universal demonstrability expected of correct principles in any established branch of science. While we may employ the currently accepted instinct analysis as a basis of illustration, it is our aim to propound methods whereby a second, more experimental and more exact phase of ergic investigation may go forward.

The new method, discussed in detail elsewhere in connection with dynamic trait measurement (7), assumes that the unitariness of a dynamic trait will show itself by coordination of strength in all behavior manifestations springing from that trait. For example, if one individual is more strongly endowed than another in the erg of escape, i.e., is of a more "timid disposition," all attitudes involving expression of timidity will *tend* to be stronger. Similarly the use of *P*-technique (6, 8) in such

<sup>13</sup> For, from a scientific, as distinct from a therapeutic, viewpoint, free association cannot be called a new method.

dynamic investigation will depend on the individual's manifestations of timidity showing a common occasion-to-occasion variation with the goal tension of this erg and will reveal the ergic pattern in terms of unique traits. In short, the unitary, independent dynamic traits, namely the ergs of fear, sex, hunger, etc., should appear as independent statistical "factors" providing inter- or intra-individual variation of drive strengths, as such, occurs. That dynamic unities will in fact reveal themselves through this method, at least as correlation clusters, has already been demonstrated by the work of E. E. Anderson (3), while the reality of ergs as independent factors is emerging from work by the present writer (9).

Examination of the underlying structure as presented in the dynamic lattice suggests that measurements of the strength of each of a large population of attitudes about many objects, as they appear at the left of the diagram, should ultimately be resolvable into the strength of a few independent ergs, as indicated at the right of the diagram. Factor analysis is only one of several tools which could be used in tracing these connections, but it is probably the best, for it alone can cope with multiply determined phenomena, where extraction of causation by experimental control of all but one influence is not possible.

It is important that this proposed investigation of dynamic structure should rest on a truly representative sample in the set of attitudes taken for the ultimate analysis. This emphasis on catholicity in the dynamic realm rests on the same arguments as those for the necessity of the "personality sphere" in general personality analysis (7). The group of attitudes—mainly to debatable social institutions—which psychologists have so far measured occupy an extremely specialized angle of the total field of attitudes.<sup>14</sup>

What remains uncertain in our hypothesis is whether other features of the dynamic lattice besides the ergic goals,<sup>15</sup> notably

<sup>14</sup> Research would clearly need to begin with a frequency-distribution study of the objects and directions of attitudes in the average member of our culture. A mere object inventory would be valueless, for obviously a great many environmental objects sustain attitudes too negligible to measure.

<sup>15</sup> Of course, if the attitude strength-measuring devices are not pure dynamic measures, uncontaminated with abilities, factors will appear also corresponding to unitary abilities, but the construction of pure dynamic measures offers no particular difficulty.



the structures which we have called total sentiments and component sentiments, will also reveal themselves by this experimental approach. Let us ask, therefore, what effect the existence of a sentiment<sup>16</sup> might be expected to have on the observed covariation of the attitudes involved in it. This amounts to examining the lattice more closely with respect to what actually happens at the points called "sentiment objects," or at the lesser sentiment objects called attitude objects.

Let us first recapitulate that any object is the focus of many component attitudes. Thus with regard to New York, you may enjoy it; you may be disgusted with it and demand that the U.S.A. secede from it; you avoid it because you find yourself invariably "broke" after you go there; or you may defend it because an enemy country is planning to attack it. Further, everybody has all of these attitudes in some degree, depending on the "stimulus situation." By the "stimulus situation," we mean the extent to which theaters and entertainment are most in consciousness, or the prices are high, or an enemy is threatening an attack. It may be objected that these physical situations do not completely define the situation but that the subject's dynamic state must be taken into account, namely, whether theaters are a hobby of his, whether his bank account is low, and whether he feels patriotic. To this we reply, as in the initial definition of attitude on p. 221, that these enter into the calculation, but not in the same part of the book. They are in the subject's attitude, not in the stimulus situation. When one operationally tests a single component attitude, say that defined by the question: "How much money would you vote to have set aside for the defense of New York?" the subject's answer will depend partly on his patriotism, an inherent aspect of his attitude strength, and partly on the "stimulus situation," i.e., the extent to which the subject perceives New York to be in danger and the extent to which he judges the loss of New York would be devastating to the nation. An attitude measurement, in short, is always in relation to a particular stimulus-

<sup>16</sup> It will be understood that since this particular experiment is couched in *R*-technique any trait will be a *common trait*. However, it could quite simply be transposed into a design in *P*-technique with a single individual, in which case the sentiment would not need to be a general one in order to manifest itself by covariation

circumstance, as our formula stated, and also is not required to take account of function fluctuation.

Similarly in any sentiment, e.g., a man's sentiment to his wife or to his bank account, there are always component sentiments to be measured, as "How much do you value your wife's preparing your meals?" "How much do you want to kiss your wife?" "How much do you propose to set aside for your wife's health insurance?" Each of these will be a function partly of the stimulus situation (not how hungry the man is, which is an inherent part of the attitude measurement, but what prospects he has of getting his meals elsewhere, or, in the last case, what threats exist to his wife's health). Each of the component attitudes or sentiments around one object may subsidiate into more than one sentiment around some farther object and into different sentiments from those served by its fellow component attitudes; for example, the attitude of avoiding New York may serve both the man's sentiment to increase his bank account and the sentiment to follow his wife's wishes. Meanwhile the attitude to defend New York may serve the different sentiment of patriotism.

Does the fact that attitudes and sentiments are composite, with components headed for different goals, mean then that there is no such thing as a single attitude or sentiment, i.e., that what we have called a *total sentiment* is an abstraction, or even a fiction, or at least something so lacking in unity as to be insusceptible to representation by a single measurement? Let us consider this in the following section with respect to sentiments, which present the attitude situation in grosser form.

### 5. Concerning the Unitary Character of a "Total Sentiment"

The question of what operational unity can be demonstrated for the verbal concept of a sentiment—which we have here distinguished more precisely as a "total sentiment" has been grossly neglected when one considers how freely the concept has been bandied about. In our present context it resolves itself into asking whether any degree of covariational unity can be expected in the strengths of the attitudes and

component sentiments which enter into the "total sentiment" about a single object. The following hypotheses, pointing to the existence of covariation, are put forward for experimental investigation.

1. Though most component sentiments about an object need have no organic relation *inter se*, they bring into existence a secondary attitude—concern for the existence and preservation of the object through which they gain their satisfactions—which will be dependent in magnitude upon the sum of the component sentiments. This is readily illustrated in the sentiment to a wife, and is perhaps more characteristic of major sentiments than of minor total attitudes. (Fears are proportional to hopes, and hates to loves.) Correlation would here be expected most strongly among the attitudes which serve the preservation and cherishing of the object (especially in fear and anger attitudes to threatening dangers) and less strongly between these and the positive component sentiment satisfactions of the sum of which the former are a function<sup>17</sup>

2. When the circumstances are such that only a single sentiment exists toward the object, i.e., component and total sentiment are the same, and the only multiplicity exists in the form of several attitudes subsidiated to the main course of action, e.g., many attitudes to objects serving the single purpose of "increasing my bank account," a correlation will be found among the attitudes. For an increase in the strength of the final activity will tend to require some increase in the strengths of all subsidiated activities.

3. While the attitude object is not the whole stimulus to the attitude, it is, as pointed out at the outset, a fairly substantial part of the situation. Insofar as different component

<sup>17</sup> Of course, this applies with appropriate reversals, whether the total satisfaction is positive or negative, i.e., where the existence of the object promotes or prevents satisfactions. This dynamic relation to the sheer existence of the object may seem to present a possible reality corresponding to the pro-con scaling which we have criticized as artificial—at best a dimension rather than a functional unit. For to ask whether a person's total ergic satisfactions are most served by the existence of a certain object or by its non-existence is identical with that involved in asking the person, "Do you like this object or do you dislike it?" However, the total sentiment measurement in fact does not provide the simple pro-con scale, for the directions (on ergic coordinates) of some kinds of liking are dynamically different from others. The manner of "liking" a wife is quite different in dynamic meaning from the manner of liking one's country or liking a church. The preservation and augmenting of the object brings a quite different composition of ergic satisfactions in its train in each case.

attitudes share encounter with the object as part of their stimulus they will covary in strength in proportion to the individual's frequency of encountering the object. (The reflexologist can conceive this as redintegration.)

The tendency for a constancy of ratio to exist among the component sentiments in a total sentiment as described in (2) and (3) above is an example of the wider principle invoked in the notion of *environmental mold trait unity* (7) affecting also cognitive and temperamental responses. The structure of the environment, social or physical, acts as a mold compelling the satisfactions, etc., of different individuals in relation to a certain institution to take on something of the same pattern. One takes the sentiment object for better or worse, accepting, if one deals with it at all, the pattern of satisfactions it provides.

A sentiment to a son, for example, must keep the expression of aggression in certain limits relative to the expression of gregariousness, else he will leave home, and gregariousness must not be overloaded with tender protectiveness else he will resent being treated as a mollycoddle. Secondly, with sentiments involved in social institutions, e.g., the attitude to a wife or to the church, the discipline of social approval will also define limits to the varieties of satisfactions. If Mr. A is far more interested in the church than is Mr. B, his interest must still have something of the same pattern as Mr. B's. It will not do for him to be inordinately interested in choir singing, while being scornfully atheistic where prayers are concerned. Co-variation of component attitudes will then arise because of the total variation: some people are more religious than others and some are more married than others.

If, therefore, a unity of covariation of parts exists in a total sentiment for the above three reasons, the term or notion of sentiment has value not merely in describing a collection of attitude structures about an object but also as defining a measurable dynamic entity.<sup>18</sup>

<sup>18</sup> The measure of a total sentiment, on this basis, would be the sum of strengths on component sentiments, weighted with regard to (not directly proportional to) the loading in the group factor formed by the total sentiment. Presumably this would be an oblique factor with direction definable in terms of ergs. It is not easy to see, however, what this measure represents in terms of behavior. There is no single course of action corresponding to the total sentiment, but only several poten-

### 6. *The Formula for an Attitude*

The above arguments from the dynamic lattice lead us to expect that the factorization of a matrix of correlations among a sufficient variety of attitudes (measured for strength) would yield (1) general factors corresponding to basic ergs and (2) narrow group factors corresponding to sentiments.<sup>10</sup>

If group factors of this kind appear, an attitude will need to be fixed in direction by coordinates corresponding to both ergs and metanergs, i.e., drives and the structures derived from drives. But this in no way affects the general principle of attitude definition. The direction of an attitude (a common, component attitude or sentiment)  $A_j$  will be defined in terms of its coordinates, as follows:

$$S_{jE_1}, S_{jE_2}, \dots, S_{jE_n}, S_{jM_1}, \dots, S_{jM_n}, S_{jA}$$

where the  $S$ 's are direction numbers (projections of the point defining the end of the vector) of the attitude vector, with respect to ergic coordinates  $E_1, E_2$ , etc., sentiment coordinates  $M_1$ , etc. (mold traits) and a specific,  $A$ , to each attitude. These  $S$ 's may be called "situational indices" since they represent, for the general population, the extent to which the situation in which the attitude operates involves (stimulates) the drives  $E_1, E_2$ , etc.<sup>20</sup>

---

tial courses of action in response to several situations which regularly present themselves. For example, a child may have an attitude of fearing a certain dog when meeting it in the street and of being angry and aggressive toward it on other occasions because it interferes with a game. In relation to *all* life situations, therefore, the attitude to the dog is one of hatred, and a single course of action which would remove the dog would represent the "total sentiment." An attitude, it has been said at the outset, is a potential course of action. But it might be desirable to distinguish between attitudes which *can* realize themselves in action and those which cannot. Total sentiments usually cannot.

This special issue—the nature of the total sentiment—and of the complex, which seemingly has the same behavioral manifestation—needs to be further theoretically pursued when experiment has been undertaken to check on the general basis of the present arguments.

<sup>10</sup> A theoretical difficulty must be faced here. In sampling attitudes blindly and at random it will necessarily happen that some will fall at different general subsidiation levels or goal distances from others. Indeed, two measured attitudes will sometimes fall in immediate sequence on a subsidiation chain, e.g., the attitude toward war and the attitude toward maintaining armaments. Such attitudes will obviously be highly correlated, will form a cluster, and will have much the same factorial composition. But the presence of such immediate subsidiations will in no way detract from the success of the factorization—unless it reaches proportions such that the factor picture is incomplete through inadequate sampling of other regions. The demonstration that two attitudes are subsidiated rather than parallel (when their factor composition is similar) would seem to require additional experiment.

<sup>20</sup> It may be objected that whereas there is only one situation,  $S_j$ , we demand separate situational indices for each dimension  $S_{jE_1}, S_{jE_2}$ , etc. This can be most

The *strength* of an attitude, which is identical with the strength of the *interest*<sup>21</sup> in that course of action, is fixed by the above vector definition only for the average person. For a given individual, *i*, the magnitude *I* of the interest strength in the attitude *A<sub>j</sub>* will be given by the following equation:

$$I_{jt} = S_{jE_1} \cdot E_{1t} + S_{jE_2} \cdot E_{2t} \dots + S_{jM_1} \cdot M_{1t} \dots + S_{jA} \cdot A_t$$

where *E<sub>1t</sub>*, *E<sub>2t</sub>*, *M<sub>1t</sub>*, etc., are the individual's endowments at that time in the dispositions *E<sub>1</sub>*, *E<sub>2</sub>*, etc., and the sentiments *M<sub>1</sub>*, etc.

### 7. *Current Practice in the Light of Ergic Measurement Theory*

Can some of the careful experiments founded on less definite or complete theory be utilized afresh in the framework of the present theory? The Thurstone school (20, 22) has from the beginning clearly thought out the notions of strength and direction as used here. But it has taken direction "as read" and has not sought to analyze it. The more pragmatic approaches of Cantril, Goodenough, Guttman, and Katz, on the other hand, have used dichotomies quite different in meaning in spite of employing similar, or the very same, terminology.

Cantril, Guttman, and others have in fact used for a single-attitude scale a number of opinion statements each of which, in our conception, is a single attitude, i.e., a single course of action, in itself. For example, "I believe Negroes should have more opportunities" (Cantril) and "Negroes should not be mistreated, but should be kept socially distinct" represent quite different courses of action toward the solution of a problem. Apparently Cantril considers that these fall upon an ergic continuum—presumably of friendliness-nonfriendliness—but it is

---

simply defended on the empirical grounds that where abilities are concerned it is found that the situation also needs a separate loading for each ability concerned, and that it is safer similarly to suppose, in the dynamic realm, that the situation has a distinct "meaning" or stimulus value for each of the ergic reactivities in the individual. A physical situation *S<sub>j</sub>* of an object-in-a-setting is fully *psychologically* defined, therefore, only when indices are known for all the basic ergs and for certain acquired sentiments. These are the "meaning" of the object in that situation, i.e., they define the vector direction of the attitude, for people in general.

<sup>21</sup> An attitude is an interest defined as to direction as well as merely to strength. If one speaks of the total interest (total sentiment) in an object, he is dealing, however, with something different from the strength of interest in a particular course of action with an object, i.e., a component attitude. The discussion of the complete relations of interest measures to attitude measures in terms of the present scheme cannot be begun here.

very doubtful whether one can be considered more friendly than the other in intention, and it is also very debatable whether one rather than the other is decidedly in the Negroes' best interest. In short, it is questionable whether these opinions form a continuum in the sense of either system (2) or (3, p. 224).

If, by luck, the test designer's intuitions about dynamic composition or social consequence actually yield opinions with

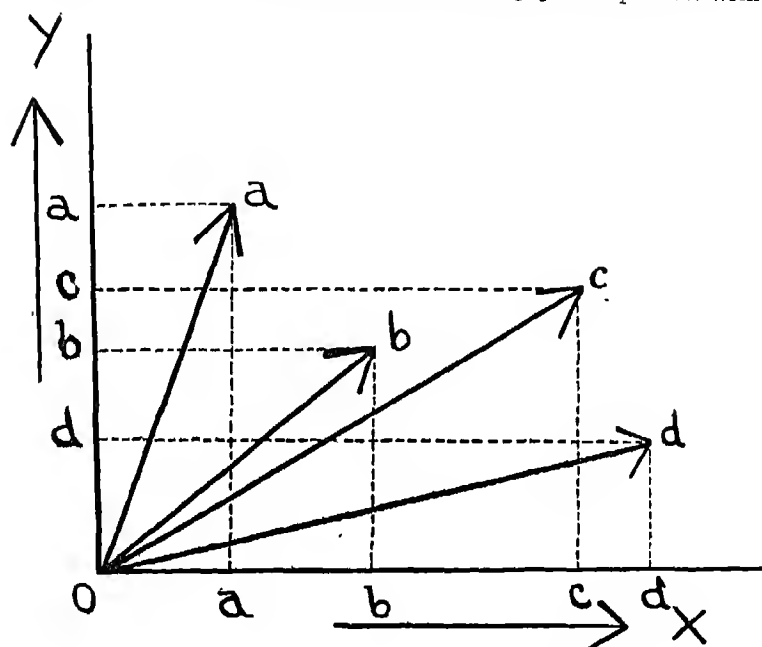


Figure II. Opinion Items in Attitude Scales

the loadings he has in mind, the situation will be as in Figure II, wherein the projections fall in a sequence on one axis  $X$  (an ergic or metanergic coordinate, as in system 3, or a social change consequence as in system 2, page 224) but in quite different sequences in other axes, such as  $Y$ . Adding these opinions will, therefore, yield an  $X$  measurement contaminated with  $Y$ , etc.

That intuition succeeds tolerably in producing some sort of meaningful continuum is shown by the fairly consistent demon-

stration of a relationship—a *U* or *J* curve—between values on this first dimension (variously conceived as direction [Cantril], extent and content [Guttman], and the second character which these experimenters ascribe to the attitude, namely, intensity [Cantril, Guttman], certainty [Guttman, F. H. Allport], or confidence [Johnson]). On the other hand, the lack of any very significant relations between attitude measures and personality traits suggests that this hidden and undefined direction continuum is more commonly a social consequence continuum and only rarely the precise dynamic coordinate demanded by the ergic theory.

Although these experimenters' use of direction may, therefore, be sometimes resolved into a clear-cut direction in the sense used here, their use of intensity or conviction resists any simple translation. Our conviction is that the second dimension is used in different senses by, for example, Cantril, F. H. Allport, Johnson, and Guttman, and again by Guttman (13)—quality, intensity—and Guttman (14)—degree of certainty, none of which has any simple relation to strength as used in this article. Here strength means interest-energy, measurable as, or validated against, work done (granted the opportunity). "Confidence," "conviction," "certainty" and perhaps "intensity of feeling," on the other hand, arise from a mixture of cognitive and dynamic operations, and even in the latter the effective variable is not simple strength, but often some complex product of conflict. I may be very certain that the sun is shining but only weakly motivated to go outside, or quite sure that a political party is rotten with graft and yet not willing to expend much energy in opposing it, or mildly convinced that whiskey is bad yet strongly motivated to get some.

Insofar as a dynamic strength enters into certainty of opinion it probably does so as a function of ego involvement only, and specifically in the need to reinforce by the self regarding sentiment and ego defense mechanisms a view in which one is opposed to the group, i.e., it is normally no indicator of real attitude strength. This obvious secondary dynamic investment origin of the *U* curve findings—from those of F. H. Allport (1) to those of Guttman (14)—was expressed years ago in one



of Chesterton's amusing paradoxes (11). "The modern habit of saying 'This is my opinion, but I may be wrong' is entirely irrational. If I say that it may be wrong, I say that it is not my opinion." Guttman's "zero point" of scaling is thus a point where the individual, because he does not have to oppose public opinion, is not required to be dogmatic and ego involved, and is therefore *unaware* of the real dynamic strength of his interests in the sense defined here. Just so, until I get asthma, I am quite unaware of my powerful dynamic interest in breathing and of the work I normally expend in connection with the attitude formula: "In normal circumstances I want *very* much to breathe air."

If words are ever to be precise symbols, respectfully used by psychologists as firm stepping stones to scientific progress, rather than as careless counterfeits confusing scientific transactions, it is necessary in attitude study to agree on some order in the use of direction, strength, intensity, certainty, conviction, content, and intensity of affect. It is suggested here that direction (content) and strength (interest) are the primary characters of an attitude, while object and stimulus-circumstance also need to be stated. Conviction, certainty, intensity of ego involvement, etc., constitute additional, secondary aspects, doubtless of considerable interest, but definable as a function of the relations among the attitude, the individual's intelligence, the security of the ego, and the individual's relation to society rather than of the attitude itself.<sup>22</sup> However, "intensity of conviction" itself requires more exploration and precise definition before numerical measurement of it can be useful and meaningful.

The immediate need of attitude research seems to be factorization of a standard, well and widely sampled set of important attitudes in the life of the normal citizen (preferably measured objectively instead of by subjective estimates of opinion) to determine the ergic coordinates for attitudes in general. Reference tests offering pure, saturated measures of these factorial coordinates could then be made available for determining the

---

<sup>22</sup> The discrepancy between intensity of conviction and the strength of the attitude as measured by performance seems likely to provide a promising measure of the *C* factor (7) of personality, involving emotionality and unrealism

direction of any given attitude in a particular, local, social or individual attitude study.

Like many problems on which researchers turn their backs in favor of more dramatic or fashionable studies, this requires a considerable effort of organizational and technical skill. But until it is made, far more energy than it would require will continue to be wasted on studies on unrelated foundations. Indeed, the methodological bankruptcy of applied attitude study, through failure to connect attitude calculus with the clinical depth psychology of sentiments, complexes, and basic ergs, has recently been amply demonstrated. Notably it has been extensively demonstrated during the war by the trivial results from attempting, by superficial suggestion methods, with million-dollar films, to influence attitudes in morale, hatred of the enemy, etc. By comparison, the undertaking here proposed to put theory and practice on a new footing is an extremely modest operation, though perhaps beyond the powers of a single researcher.

### 8. Summary

1. An attitude is a vector definable by strength, direction, object (or object situation), and stimulus situation.

2. *Direction* is definable by reference to a coordinate system of primary ergic goals (and possibly of environmentally molded sentiments).

3. If the hypothesis about ergs and ergic structure propounded here and elsewhere (6, 7) is correct, these ergic coordinates can be discovered by a factor analysis of a sample of the total attitudes of the individual. When founded on *R*- or *Q*-technique this will give common coordinates; but on *P*-technique the coordinates of individual sentiments and drives. Ergs are likely to be practically general factors; sentiments, comparatively narrow group factors.

4. Attitudes fall in subsidiation sequences, ending in ergic goals, constituting a typical form which we have called the "dynamic lattice." The dynamic lattice is a reticulum formed by the interflux and subsidiation of attitudes.

5. Examination of the dynamic lattice shows that we must distinguish between two dynamic structures: component atti-

tudes or sentiments which express a single course of action in response to a single situation or set of situations and total sentiments which are structures formed by several courses of action (from several recurring situations). The latter do not have any single course of action corresponding to the theoretical dynamic resultant of the component attitudes, but they have a dynamic unity in that dynamic trends have to be centered on the preservation of the object from which the diverse component satisfactions are obtained.

6. With complex, abstract objects—notably social institutions—there is little one can do except to preserve or to destroy. This possibility of only one course of action, as well as the unitariness of drive mentioned in the last part of 5, above, gives a certain basis for the current practice of making pro-con scales. But this is an approximate procedure because the approval-disapproval has a widely different dynamic composition (direction) in each case.

7. The *strength* of an attitude is measurable by the interest and activity in the course of action concerned (or by such indirect manifestations as prove to be valid representatives of this). Intensity of conviction is a far more complex, secondary variable. Strength is derivable from the formula:

$$I_{jt} = S_{jB_1} \cdot E_{1t} + S_{jB_2} \cdot E_{2t} \dots + S_{jM_1} \cdot M_{1t} \dots + S_{jA} \cdot A_t$$

in which  $S_{jB_1}$ ,  $S_{jB_2}$ , etc., are direction cosines (unit projections) defining the *direction* with reference to the ergs which constitute the primary dynamic coordinates and  $E_{1t}$ ,  $E_{2t}$ , etc., are the individual *i*'s endowments in the dispositions  $E_1$ ,  $E_2$ , etc.

8. Corollaries from the ergic theory would lead to a calculus of conflict and adjustment as well as to laws concerning the relation of social emergents to individual dynamics.

## REFERENCES

1. Allport, F. H. and Hartmann, D. A. "The Measurement of Motivation of Atypical Opinion in a Certain Group." *American Political Science Review*, XIX (1925), 735-760.
2. Allport, G. W. *Personality: A Psychological Interpretation*. New York: Henry Holt & Co., 1937.
3. Anderson, E. E. "The i . . . . . Drives in the Male Albino Rat: III. Among Measures of Emotional, Sexual,

- and Exploratory Behavior." *Journal of Genetic Psychology*, LIII (1938), 335-352.
4. Cantril, H. "The Intensity of Attitude." *Journal of Abnormal and Social Psychology*, XLI (1946), 129-136.
5. Cattell, R. B. *A Guide to Mental Testing*. London: University of London Press, 1936.
6. Cattell, R. B. "Sentiment or Attitude: The Core of a Terminological Problem in Personality Research." *Character and Personality*, IX (1940), 6-18.
7. Cattell, R. B. *The Description and Measurement of Personality*. Yonkers-on-the-Hudson: World Book Company, 1946.
8. Cattell, R. B. "The Dimensions of Change of Culture Patterns." To be published.
9. Cattell, R. B. "The Ergic Structure of the Rat." In preparation.
10. Cattell, R. B. and Luborsky, L. B. "P-technique Demonstrated in Determining Psycho-physiological Source Traits in a Normal Individual." *Psychometrika* (In Press).
11. Chesterton, G. K. Preface to *The Book of Job*. London: S. Wellwood, 1907.
12. Goodenough, W. H. "A Technique for Scale Analysis." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, IV (1944), 179-190.
13. Guttman, L. "A Basis for Scaling Qualitative Data." *American Sociological Review*, IX (1944), 139-150.
14. Guttman, L. "Experiments on the Measurement of the Intensity Function and Zero Point in Attitude Analysis." Headquarters, Army Service Forces, Information and Education Division, Report No. D-1, 1945.
15. Johnson, D. M. "Confidence and the Expression of Opinion." *Journal of Social Psychology*, XII (1940), 213-220.
16. Katz, D. *The Measurement of Intensity in Gauging Public Opinion*. Princeton: Princeton University Press, 1944.
17. McDougall, W. *The Energies of Man: A Study of the Fundamentals of Dynamic Psychology*. New York: Charles Scribner and Sons, 1933.
18. McNemar, Q. "Opinion-Attitude Methodology." *Psychological Bulletin*, XLIII (1946), 289-375.
19. Murray, H. A. et al. *Explorations in Personality*. Cambridge: Harvard University Press, 1938.
20. Thurstone, L. L. "The Theory of Attitude Measurement." *Psychological Review*, XXXVI (1929), 221-241.
21. Thurstone, L. L. "The Measurement of Change in Social Attitude." *Journal of Social Psychology*, II (1931), 230-235.
22. Thurstone, L. L. and Chave, E. J. *The Measurement of Attitude*. Chicago: University of Chicago Press, 1929.

# THE CORNELL TECHNIQUE FOR SCALE AND INTENSITY ANALYSIS<sup>1</sup>

LOUIS GUTTMAN

Cornell University

## I. *Introduction*

DURING the course of the war a new approach to the problem of scaling attitudes and public opinion, called *scalogram* analysis, was developed by the writer to aid in the study of the morale and related aspects of the United States Army. This approach has wide ramifications not only for attitude and opinion research, but for many other fields like market research, mental testing, and elsewhere where it is desired to quantify qualitative data. Not much has yet been published<sup>2</sup> on this approach during the five years it has been used by the Army, so that it has not been readily available to other research workers.

The work of the Research Branch of the Army Service Forces, done under the scientific leadership of Professor Samuel A. Stouffer, will be described in several volumes now being completed. One of these volumes contains a rather comprehensive treatise on the theory and practice of scalogram analysis as carried out by the Research Branch.

The purpose of the present paper is to describe another technique for scalogram analysis which can be used immediately by research workers. Justification for the technique follows from the general theory and evidence to be published in the forthcoming volumes on the Research Branch. We shall call it the *Cornell technique* for scalogram analysis to distinguish it from several alternative devices, since it was developed first for teaching purposes at Cornell. It is hoped that the reader may be

<sup>1</sup> A paper presented to the Conference on Measurement of Consumer Interest, University of Pennsylvania, May 17-18, 1946

<sup>2</sup> The basic concepts are available in Louis Guttman, "A Basis for Scaling Qualitative Data," *American Sociological Review*, IX (1944), 139-150.

able to master the technique from this present description. For a fuller exposition of the theory and a discussion of the problems of reliability, validity, and the like, he is referred to the forthcoming books on the work of the Research Branch.

*The Scalogram Analysis Approach.*—The Cornell technique is a procedure for testing the hypothesis that a universe of qualitative data is a scale for a given population of people, using the scalogram approach. It may also be used to test the hypothesis that the data form a quasi-scale. Of the several techniques now available for scalogram analysis,<sup>3</sup> the one to be described here seems to be among the simplest and most convenient for general use. It requires no special equipment and involves only very simple clerical procedures which can readily be carried out by persons unskilled in statistics.

The various techniques just referred to all do the same job since they follow the same scalogram theory; they differ only in how the work is arranged. The initial steps are common to all. First, the universe of content to be studied is defined. In an attitude or opinion study, this means deciding on the general content of the questions to be asked. Second, the population of people is defined. In an attitude or opinion survey, this means that the class of people to be interviewed is delimited.

Next come two kinds of sampling problems. One kind is the ordinary problem of random sampling of people, and the other is the sampling of items. For these two sampling problems, it is helpful to distinguish between the pre-test stage of a study and the final survey. Many fewer people can be used in a pre-test than must be used in the final survey, but fewer items can be used in the final survey than must be used in the pre-test.

<sup>3</sup> The first technique employed laborious least squares computations. See Louis Guttman, "The Quantification of a Class of Attributes: A Theory and Method of Scale Construction" in P. Horst et al., *The Prediction of Personal Adjustment*, Social Science Research Council, 1941, pp. 319-348. The standard procedure used by the Research Branch involves the use of scalogram boards especially invented for this purpose by the writer; these boards are simple to build and to operate, and a description of them will be in the forthcoming publication. A tabulation technique has been devised by another member of the Research Branch; see Ward H. Goodenough, "A Technique for Scale Analysis," *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, IV (1944), 179-190. The Cornell technique was devised by the writer at first for teaching purposes, and has proved to be very useful for general research purposes. A brief statement of the procedure as carried out on IBM equipment has already been noted in E. William Noland, "Worker Attitude and Industrial Absenteeism: A Statistical Approach," *American Sociological Review*, X (1945), 503-510.

In the pre-test for a survey, about 100 persons will usually constitute an adequate sample of the population to test the hypothesis of scalability. If the hypothesis is accepted, the items can then be used in the final study of the usual 3,000 or so people to obtain reliable proportions at each scale rank.

The other sampling problem is of quite a different nature; it consists of sampling the universe of content. In an attitude or opinion survey, this is done by constructing some questions which contain the required general content. In a pre-test, about a dozen questions usually can constitute an adequate sampling of the content. Since questions are constructed by the research workers, they do not fall into any standard random sampling scheme, and standard random sampling theory does not apply here. Instead, it is shown by the theory of scale analysis that *almost any* sample of about a dozen questions from the universe is adequate to test the hypothesis that the universe is scalable, provided the range of content desired is covered by the questions. If the hypothesis is accepted that the universe is scalable, then fewer questions can be used in the final study if fewer ranks are actually needed for the purposes of the final research.

Having defined the universe of content and the population of people, and having drawn a sample from each, the fifth step is to observe each person in the sample on each item or question in the sample. In an attitude or opinion survey where a questionnaire is used, this involves having the people indicate their answers to each question of the questionnaire.

*The Hypothesis of Scalability.*—The problem now is to test the hypothesis, on the basis of the pre-test sample data, that the entire universe of items forms a scale for the entire population of people. Let us review what this hypothesis implies in order to see what the technique of analysis is trying to do.

The universe is said to be scalable for the population if it is possible to rank the people from high to low in such a fashion that from a person's rank alone we can reproduce his response to each of the items in a simple fashion.<sup>4</sup> It is understood that a perfect scale is not to be expected in practice. Data have

<sup>4</sup> For a basic discussion of the theory of scales, see Louis Guttman, "A Basis for Scaling Qualitative Data," *ibid.*

been considered sufficiently scalable if they are about 90 per cent reproducible, and if certain other conditions (to be explained later) are satisfied. For clarity, though, let us consider first a hypothetical perfect scale.

Suppose that a question from the universe is asked of a population concerning a certain political issue and that the responses are as follows:

Agree	60%
Undecided	10
Disagree	30
<hr/>	
	100%

If "Agree" means a more favorable opinion than "Undecided," if "Undecided" is more favorable than "Disagree," and if the universe is perfectly scalable, then the following must be true. The highest 60 per cent of the people must be those who said "Agree"; the next highest 10 per cent must be those who said "Undecided"; and the lowest 30 per cent must be those who "Disagree." If another question from this scalable universe is asked and the responses are 20 per cent "Yes" and 80 per cent "No," and if "Yes" means a more favorable attitude than "No," then the top 20 per cent of the people must be those who said "Yes" and the bottom 80 per cent must be those who said "No." From the rank of a person, we can now deduce what his response must be to each of these two questions. Any person in the top 20 per cent of the population must have said "Agree" to the first question and "Yes" to the second question. Any person lower than the top 20 per cent but not lower than the top 60 per cent said "Agree" to the first question and "No" to the second question. Any person below the top 60 per cent but not below the top 70 per cent said "Undecided" to the first question and "No" to the second, and the rest of the people, the bottom 30 per cent, said "Disagree" to the first question and "No" to the second.

The various techniques for scalogram analysis are devices to find the rank order for the people which will best reproduce their responses to each of the items in this fashion. If the universe were a perfect scale, all of the techniques would involve little



work and there would not be much to choose between them. It is the presence of imperfect reproducibility that raises the problem of technique.

The Cornell technique works by successive approximations. Usually just two approximations suffice to reject or accept the hypothesis of scalability. A first trial rank order for the people is established by a simple scoring scheme. For illustrative purposes, let us work out an actual case in detail. This illustration is not to be taken as a model of perfect research, but rather only to provide an example of the steps to be followed.

*An Example of the Cornell Technique.*—It was desired to find out if the students in a certain class in race relations had a scalable attitude toward one of their textbooks, *A Nation of Nations*, by Louis Adamic. A questionnaire with seven questions was made out and administered to the class of 50 students. Both the number of questions and the number of students were smaller than those ordinarily used in a pre-test; they were used here only because these smaller numbers permit displaying the full data.

The seven questions were as follows:

### *A Nation of Nations*

#### Questions

1. *A Nation of Nations* does a good job of analyzing the ethnic groups in this country.

Strongly agree	4	Agree	3	Undecided	2
Disagree	1	Strongly disagree	0		

2. On the whole, *A Nation of Nations* is not as good as most college textbooks.

Strongly agree	0	Agree	1	Undecided	2
Disagree	3	Strongly disagree	4		

3. Adamic organizes and presents his material very well.

Strongly agree	4	Agree	3	Undecided	2
Disagree	1	Strongly disagree	0		

4. As a sociological treatise, Adamic's book does not rate very high.

Strongly agree	0	Agree	1	Undecided	2
Disagree	3	Strongly disagree	4		

5. Adamic does not discuss any one group in sufficient detail so that a student can obtain a real insight into problems of ethnic group relations in this country.

Strongly agree	0	Agree	1	Undecided	2
Disagree	3	Strongly disagree	4		

6. By providing a panorama of various groups, *A Nation of Nations* lets the student get a good perspective on ethnic group relations in this country.

Strongly agree	4	Agree	3	Undecided	2
Disagree	1	Strongly disagree	0		

7. *A Nation of Nations* is good enough to be kept as a textbook for this course.

Strongly agree	4	Agree	3	Undecided	2
Disagree	1	Strongly disagree	0		

## II. Content Scale Analysis

We now describe, step by step, how the analysis of the responses is carried out by the Cornell technique:

1. Weights for the first trial are assigned to each category of each question, using the successive integers beginning with zero. In this example, since each set of answers has five categories, the weights range from 0 to 4. In each question, the higher weights are assigned to the categories judged to express a more favorable attitude. This judging of ranks of categories is not to be regarded as final; the consequent analysis will either verify the judging or determine how to revise it.

2. A total score is obtained for each person by adding up the weights of the categories he falls into. In our example, since the maximum weight for each person is four, and the total number of questions is seven, the total scores can range from zero to 28.

3. The questionnaires are shuffled into rank order according to the total scores. In our example, we have arranged them from high to low.

4. A table is prepared, like Table 1 below, with one column for each category of each question and one row for each person. Since each of our questions has five categories, and since there are seven questions, we have 35 columns in our table. There are 50 students, so we have 50 rows. The first five columns are for the five categories of the first question, the second five columns for the five categories of the second question, etc.

5. The response of each person to each question is indicated on the table by placing an X in his row in the column for each category into which he falls. In our example, we have labeled the columns according to the questions and the weights of the categories. The first person is the one with the highest score, which is 28. He had checked the response weighted 4 in each of the questions, so he has seven X's in his row, each under the respective columns for the categories with weight 4. There were two persons with a score of 25. The arrangement of people with the same score is arbitrary. Of the two persons in our example with a score of 25, the one placed first had a response of 4 to the first two questions, a response of 3 to the third question, of 4 to the fourth question, of 3 to the fifth and sixth questions, and of 4 to the seventh question. Similarly, the X's in Table 1 indicate the response of each of the remaining persons to each question. Every person answers every question<sup>5</sup> so that there are seven X's in each row. *Table 1 gives a complete record of all the data obtained by the survey with respect to this area.*

6. At the bottom of Table 1 are the frequencies of response for each category. Category 4 of question 1 had nine people in it, whereas category 3 of the same question had 27 people, etc. The sum of the frequencies of the five categories in each question is always the total number of people in the sample, which in this case is 50.

7. Now we come to the test for scalability. If the universe

---

<sup>5</sup> If people sometimes fail to respond to a question, then another category is added entitled "No Answer," which is weighted and treated like any other category for that question. In the present example, there were no "No Answers."



TABLE 1—(Continued)

Score	1			2			3			4			5			6			7																
	4	3	2	1	0	4	3	2	1	0	4	3	2	1	0	4	3	2	1	0	4	3	2	1	0										
16	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
16	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
15	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
15	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
15	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
14	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
14	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
13	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
13	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
12	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
12	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
11	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
11	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
10	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
9	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
8	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
7	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
7	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
7	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
6	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
4	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
Freq	9	27	2	12	0	8	24	0	13	5	10	25	8	7	0	3	7	16	14	10	3	14	5	21	7	9	21	7	12	1	11	19	5	11	4

is a scale and if the order in which we have placed the people is the scale rank order, then the pattern of X's in Table 1 must be of a particularly simple kind. Let us consider the first question in the Table. If response 4 is higher than response 3, and if 3 is higher than 2, and if 2 is higher than 1 (response 0 happens to have no frequency in this case), then the nine people in category 4 should be the top nine people. Actually, six of them are the top six and the other three scatter farther down the column. Similarly, the twenty-seven people in category 3 should be below the first nine people and should go down to the thirty-sixth person ( $36 = 9 + 27$ ). Again, this is not perfectly true for our data. A similar examination for the other items shows that there is a substantial error of reproducibility in their present form. The approximate number of errors need not be counted at this stage, since it is evidently more than 15 per cent of all the 350 responses ( $350 \cdot .7 > 50$ , the number of questions times the number of people) in Table 1.

8. It has seldom been found that an item with four or five categories will be sufficiently reproducible if the categories are regarded as distinct. One reason for this is the verbal habits of people. Some people may say "Strongly Agree" where others may say "Agree," whereas they have essentially the same position on the basic continuum but differ on an extraneous factor of verbal habits. By combining categories, minor extraneous variables of this kind can be minimized. By examining the overlapping of the X's within the columns of each question, it can be determined how best to combine the categories so as to minimize the error of reproducibility for the combinations. In question 2, for example, categories 4 and 3 seem to intertwine, so they are combined. Similarly, in the same question, categories 1 and 0 seem to intertwine, so they are combined. In question 4, on the other hand, we combine categories 3, 2, and 1, leaving categories 4 and 0 separate. The way to combine categories is determined for each question separately. The combinations decided upon for this example on the basis of Table 1 are given in Table 2.

If it is desired to keep many scale types, then as little combination as possible should be done. However, if not many

TABLE 2  
*Combinations of Categories*

Question	Combinations
1 .....	(4) (3) (2,1,0)
2 .....	(4,3) (2,1,0)
3 .....	(4,3,2) (1,0)
4 .....	(4) (3,2,1) (0)
5 .....	(4,3,2) (1,0)
6 .....	(4,3) (2,1,0)
7 .....	(4) (3) (2,1,0)

scale types are desired, the categories may be combined as far as one wishes even though this may not raise reproducibility. There is no harm in combining categories that could otherwise remain distinct with respect to scale error; all that is lost by such a combination is one scale type. On the other hand, categories may *require* combination in order to reduce error; they should be combined in the manner indicated by Table 1 and not arbitrarily.

9. A second trial rank order for the people can now be established on the basis of the combined categories. This is done by reassigning weights. Since the first question now has three categories (that is, three combinations), these are assigned the weights 0, 1, and 2. Question 2 now has two categories. These could be assigned the weights 0 and 1. In the present example the weights 0 and 2 are used instead, since keeping the range of weights relatively constant from item to item often helps to establish a better ranking for the people when there is error of reproducibility present <sup>6</sup>

10. Each person is now given a new score which represents his second trial rank order. This is done by re-scoring his questionnaire according to the new weights. This re-scoring is easily done from Table 1. Using a strip of paper which is as wide as the Table, the new weights for the old categories can be written directly on the edge of the strip. Placing the strip across the row for a person, the weights are added according to where the X's lie. For our example, the strip would have for its first five columns the weights 2, 1, 0, 0, 0, weight 2 being placed in the column which was the old category 4, the weight 1 in the column which was the old category 3, and the 0's being

<sup>6</sup> In a perfect scale, *any* set of weights, provided they have the proper rank order for the categories, will yield a perfect rank ordering for the people.

in the old columns 2, 1, and 0 which are now combined. For question 2, the strip would have for the five columns the weights 2, 2, 0, 0, 0. Similarly, the new weights for the other questions can be written down to be used over the old columns of Table 1. The person who was formerly first on Table 1, with a score of 28, now has a score of  $2 + 2 + 2 + 2 + 2 + 2 + 2 = 14$ . The second person in Table 1 also gets a score of 14. The third person in Table 1 now gets a score of  $2 + 2 + 2 + 1 + 2 + 2 + 2 = 13$ ; and so on for each person.

11. The people are now shifted into the rank order of their new scores, and Table 3 is prepared from the combined data just as Table 1 was prepared from the original data. Question 1 now has three columns, question 2 has two columns, etc. The data of Table 1 are modified to fit Table 3 according to the combinations indicated in Table 2. The columns of Table 3 now refer to the combined categories, and the scores of Table 3 are the second trial scores just obtained in the preceding step.

12. The error of reproducibility in Table 3 seems much smaller than in Table 1, and we shall now count up the actual errors. This is done by establishing *cutting points* in the rank order of the people which separate them according to the categories in which they would fall if the scale were perfect. For question 1, which has three categories, we need two cutting points. The first seems to fall between the last person with score 12 and the first person with score 11. All people above this cutting point should be in category 2, and all people below should not be in category 2. Since there is one person in category 2 below this point, we have one error for category 2. A second cutting point is needed to separate category 1 from category 0; since these two categories overlap somewhat, its exact location is not essential since moving it slightly up or down will not change the amount of error. It should be placed so as to minimize the error, but this may be done in several adjacent ways. One way is to place the cutting point between the second and third persons with score 4. Below this point we find three errors in category 1, and above this, we find five errors in category 0. The total number of errors in question 1 is  $1 + 3 + 5 = 9$ . Since there are 50 responses to question 1,



TABLE 3  
A NATION OF NATIONS  
Second Trial Content

Score	1			2		3		4			5		6		7		
	2	1	0	2	0	2	0	2	1	0	2	0	2	0	2	1	0
14	x			x		x		x			x		x		x		
14	x			x		x		x			x		x		x		
13	x			x		x		x	x		x		x		x		
13	x			x		x		x			x		x		x		
13	x			x		x		x			x		x		x		
13	x			x		x		x			x		x		x		
12	x			x		x		x			x		x		x		
12	x			x		x		x			x		x		x		
11		x		x		x		x			x		x			x	
11		x		x		x		x			x		x			x	
11		x		x		x		x			x		x			x	
11			x	x		x		x			x		x		x		
11		x		x		x		x			x		x			x	
11				x		x		x			x		x			x	
11		x		x		x		x			x		x			x	
11		x		x		x		x			x		x			x	
11		x		x		x		x			x		x			x	
11		x		x		x		x			x		x			x	
10		x		x		x		x			x		x		x		
10		x		x		x		x			x		x			x	
10		x		x		x		x			x		x		x		
9	x				x	x		x			x		x		x		
9		x		x		x		x			x		x			x	
9		x		x		x		x			x		x			x	
9		x		x		x		x			x		x			x	
9		x		x		x		x			x		x			x	
9		x		x		x		x			x		x			x	
8		x		x		x		x			x		x			x	
7		x		x		x		x			x		x			x	
7		x		x		x		x			x		x			x	
7		x		x		x		x			x		x			x	
7			x		x	x			x		x		x			x	
6			x		x		x	x			x		x			x	
6		x		x		x		x			x		x			x	
6		x		x		x		x			x		x			x	
6			x		x	x		x			x		x		x		
5			x		x	x		x			x		x			x	
4		x			x	x		x			x		x			x	
4		x			x		x	x			x		x			x	
4			x		x	x			x		x		x			x	
3		x			x	x			x		x		x			x	
3		x			x		x		x		x		x			x	
3			x		x	x			x		x		x			x	
2			x		x	x			x		x		x			x	
2			x		x	x			x		x		x			x	
2			x		x	x			x		x		x			x	
2			x		x	x			x		x		x			x	
1		x			x		x			x		x		x			x
1			x		x		x		x			x		x			x
1			x		x		x			x			x				x
0			x		x		x		x			x		x			x
Freq.	9 27 14			32 18		43 7		3 37 10			22 28		30 20		11 19 20		

this means 18 per cent error. This error could be reduced, of course, by combining the last two columns and leaving question 1 as a dichotomy. Then there would be only the one error in the first column. Such a further dichotomization need not be done if there is relatively little error in the other questions so that the error over all questions is not much more than 10 per cent.

Question 2 has two categories in the second trial, and the cutting point which will minimize the error is between the last two scores 6, which makes two errors in the first column and four errors in the second column of question 2. Similarly, question 3 has a cutting point between the last score 2 and the first score 1, leaving three errors in its second column. Question 4 gets two cutting points, questions 5 and 6 one cutting point, and question 7 two cutting points. The total number of errors in the whole of Table 3 is 40, which is 11 per cent of all the responses. We can, therefore, conclude in view of the fact that much of the error occurs in question 1 and could be eliminated by combining two categories in that question, that this area is scalable. From a person's rank order, we can reproduce his response to each question *in terms of combined categories* with 89 per cent accuracy (or better, if we combine the last two columns of question 1).

13. The per cent reproducibility alone is not sufficient to lead to the conclusion that the universe of content is scalable. The frequency of responses to each separate item must also be taken into account for a very simple reason. Reproducibility can be artificially high simply because one category in each item has a very high frequency. It can be proved that the reproducibility of an item can never be less than the largest frequency of its categories, regardless of whether the area is scalable or not. For example, question 3 in Table 3 has quite an extreme kind of distribution. Forty-three students are in one category, and seven in the other. Under no circumstances, then, could there be more than seven errors made on this item, regardless of whether or not a scale pattern existed. Or again, question 4 in Table 3 has thirty-seven cases in its modal category and thirteen cases in the other two categories. Under no circum-

stances, then, could item 4 have more than thirteen errors. Clearly, the more evenly the frequencies are distributed over the categories of a given item, the harder it is for reproducibility to be spuriously high. Questions 5 and 6 in Table 3 each have high reproducibility, each having five errors; these are not artificially high because question 5 has only twenty-eight cases in its more frequent category and question 6 has thirty cases for its modal frequency. The maximum possible error for question 5 is twenty-two, and for question 6 it is twenty. The scale pattern represents quite a substantial reduction from this maximum error. An empirical rule for judging the spuriousness of scale reproducibility has been adopted to be the following: no category should have more error in it than non-error. Thus, the category with weight 2 in question 1 (Table 3) has eight non-errors and one error; category with weight 1 in this same question has twenty-four non-errors and three errors; category 0 has nine non-errors and five errors. Thus question 1 fits this rule. Question 3 comes perilously near to not fitting the rule. While the first column of question 3 (in Table 3) has no error, the second column has three errors compared to four non-errors. Similarly, the first column of question 4 has one error compared to two non-errors. It is because evenly distributed questions like 5 and 6 have little error and because the errors in the other questions, like in 3 and 4, are not too widely displaced from where they ought to be, that we consider this area to be scalable.

In constructing a sample of items to be used in a test for scalability, at least some of the items should be constructed, if at all possible, to obtain a uniform distribution of frequencies. Such items afford a good test of scalability. However, items with non-uniform frequencies are also needed in order to get differentiated scale types, so both kinds of items must be used. The more categories that are retained in an item, the sharper is the test for scalability, because error—if it really should be there—has a better possibility to appear when there are more categories.

### III. *Intensity Analysis*

*Separating "Favorable" from "Unfavorable" People.*—Since the expression of opinion about the textbook, *A Nation of Na-*

tions, is sufficiently scalable, it is meaningful to say that one student likes the book better than another. There is a meaningful rank ordering of the students according to their opinion of the book. This ordering is expressed by the scale scores assigned in the second trial. A student with a higher score than another says the same or better things about the book (within scale error).

There is a further question that is of interest to the research worker. Given that the individuals can be ranked according to their degree of favorableness, is there a cutting point in this rank order such that we can say that all people to the right of the point are "favorable" and all people to the left are "unfavorable"? One person may be more favorable than another, yet both may be favorable. Obtaining just a rank order does not distinguish between being favorable and being unfavorable; it merely reflects being *more* favorable and *less* favorable and does not tell if a point is reached beyond which being *less* favorable actually means being "unfavorable."

An objective answer to this problem is provided by the use of the *intensity function*.

The theory of intensity analysis will be explained in detail in the forthcoming publication on the work of the Research Branch. For our purposes, all we need to know is that it provides a solution to the traditional problem of question "bias." No matter how questions are worded or "loaded," use of the intensity function will yield the same proportion of the group as favorable and unfavorable. The intensity function provides an invariant zero point for attitudes and opinions.

There are several techniques for obtaining intensity in a questionnaire, as will be discussed in the volumes to be published on the work of the Research Branch. We shall discuss only two here, as carried out by the Cornell technique. These are very simple indeed to perform. The first is the *fold-over* technique, and the second is the *two-part* technique. The fold-over technique is theoretically less justifiable than the two-part technique. However, it does have some practical advantages in some cases.

*The Fold-Over Technique.*—The fold-over technique con-

sists simply of re-scoring the content questions in order to obtain an intensity score. This is easily done for the form of question used to study opinions about *A Nation of Nations*. The following weights are assigned to the check list of answers: "Strongly agree" and "Strongly disagree" receive a weight of 2; "Agree" and "Disagree" receive a weight of 1; and "Undecided" receives a weight of 0.<sup>7</sup> Thus the apparently more intense responses receive higher weights, and the apparently less intense responses receive lower weights, regardless of whether the responses appear to be "favorable" or "unfavorable."

Weighting the responses in this way means that in order to obtain an intensity score, we are in fact combining opposite ends of the check list, so that there are but three (combined) intensity categories per question. Intensity, as obtained in this fashion, is not in general scalable. Instead, it forms what is called a quasi-scale. In a quasi-scale, there is no perfect relationship between a person's response to each question and his score on all the questions; instead, there is a gradient. The higher a person's score, the more *likely* he is to give a high response to each item, but there is not the high certainty that exists in the case of a scale. This can be seen in our example of Adamic's textbook. Arranging the data into a scalogram according to total intensity score, we obtain the configuration shown in Table 4. Each question now has three categories which represent the three intensity steps. There is a density gradient of responses. There are no clear-cut streaks in the category columns but, instead, gradually tapering densities that blend from one category into the next. Combining categories still will not yield a scalable pattern.

According to the basic theory of intensity analysis, intensity should be a perfectly scalable variable. The equations of scale analysis show that there is a second component in every scale of content which is a U- or J-shaped function of the scale scores. This component has been identified as the intensity function of the content scale. What we are trying to do is to obtain this intensity by direct empirical methods. The fact that our ob-

<sup>7</sup> These weights can be written on a strip of paper to be put over Table 1 and added up there to obtain an intensity score for each person.



TABLE 4—(Continued)

[illegible]

served intensity is not perfectly scalable shows that it is not the pure intrinsic intensity we are seeking. No perfect way has yet been found for obtaining intensity, but satisfactory results are obtainable even with imperfect intensity techniques. Instead of a perfect intensity function, we will get one that can have considerable error in its relationship to the content scale scores.

*Plotting Intensity Against Content.* The empirical intensity function is obtained by plotting the intensity scores just obtained against the content scores obtained from the previous section from the second content trial. The scattergram is

TABLE 5  
A NATION OF NATIONS  
*Scattergram of Intensity and Content*

Intensity	Content (Second Trial)							Total
	0-2	3-5	6-8	9-10	11	12-13	14	
14							1	1
13								0
12	1							1
11				1		2	1	4
10					1	2		3
9	4	1	1		1	1		8
8	2	1	1	2				6
7	1	1	4	2	4	1		13
6		1	3	4	2			10
5		1						1
4		1			1			2
3		1						1
Total	8	7	9	9	9	6	2	50

shown in Table 5. The frequency in boldface in each column of Table 5 corresponds to the position of the median intensity for the respective columns. If the pure intrinsic intensity were being measured by our technique, there would be no scatter about these medians at all, but intensity would be a perfect U- or J-shaped function of the content scores. Despite the presence of error, however, the approximate shape of the true intensity function is clear from the shape of the curve along which the columnar medians lie. The curve descends from the right, or the more favorable content scores, reaches its low point at the next to the last interval to the left (contents scores 3-5), and then rises again at the last interval to the left. The



content scores 3-5, then, must be the approximate interval which contains the zero-point of the attitude. Students to the left of this interval can be said to have *negative* attitudes and students to the right can be said to have *positive* attitudes toward the textbook. Students in the 3-5 interval cannot be divided into positives and negatives without the aid of additional questions which will help to differentiate more precisely between their ranks.

On the basis of Table 5, we can conclude, then, that about 8 students did not like the textbook, 35 students did like the textbook, while 7 students were in between these. This division of the students into those with favorable and those with unfavorable attitudes does not depend upon the particular way we worded our questions. The same intensity curve, with the same proportion to the right and to the left of the zero-point, would have been obtained if we had used other questions or other wordings, provided only that these other questions were scalable with the present questions. Proof of this invariant property of the intensity function is given in the forthcoming volumes on the Research Branch's work.

*Need for Larger Sample of People.*—An important caution must be sounded here. The example we are working with must be regarded as a highly fortunate one in one sense for the purposes of this exposition. It is rare indeed to find as low error as we have in the intensity function so that the intensity curve and zero-point show out quite clearly on the basis of our small sample of 50 cases. In general this will be far from the case. To perform an intensity analysis safely, when there is a substantial error present—which is the usual case—ordinarily from one to three thousand cases are needed to obtain stable medians. To perform the scalogram analysis, it is also safer to use more than 50 cases. A hundred cases is a desirable minimum to use in the pre-test, as well as a dozen or so items instead of seven as we have used in our illustrative example. If the pre-test has established that the universe of items is scalable, the final study should be done on the usual number of cases used in opinion surveys if reliable results with respect to intensity are to be obtained. The hypothesis of scalability can be tested in

a pre-test on relatively few people because of its specialized character. However, *proportions* of the population at any given rank or on one side of the zero-point are subject to ordinary sampling error; larger samples of people must be used for reliable results with respect to them.

*Drawbacks to the Fold-Over Technique.*—The fold-over technique for intensity has two theoretical drawbacks to it, as well as some practical ones. First, the intensity scores obtained thereby are not experimentally independent of the content scores because exactly the same answers are used for both of the scores. This may give rise to some spuriousness in the relationship between the two. Second, it assumes that "Strongly agree" and "Strongly disagree" are approximately equal in intensity and opposite in direction, and similarly for "Agree" and "Disagree," while it is assumed that "Undecided" approximately straddles the zero-point. These assumptions need not be true at all. In fact, the occasional falsity of these assumptions is one contribution to error in the obtained intensity scores.

If the assumptions were true, life would be much easier for research workers. It would not be necessary to ask a series of questions in order to obtain a zero interval because the "Undecided" category for any question would provide such an interval. But, unfortunately, it is clear that in a series of questions on the same issue, the people who are "Undecided" on one question can all be "Agreed" on another question. It is just because we cannot interpret the bias of a question by looking at its content that such a technique like that of the intensity function is needed.

While the fold-over technique does have these two theoretical drawbacks, it does seem to average out the errors involved in violating the above assumptions and to provide a proper U-or J-shaped curve in many cases.

A practical disadvantage to the fold-over technique has been found in the case of man-in-the-street interviews, where people would avoid the "strongly" categories almost completely, so that not much differentiation in intensity could be obtained. In such a case, a two-part technique is necessary. An advantage of the fold-over over the two-part technique is that it takes less

space and time in administering questionnaires. The two-part technique will be illustrated in the next example.

#### IV. Another Example of Content and Intensity Analysis

*A Universe Is Not Necessarily a Scale.*—A set of items constructed from a single universe of content is not necessarily scalable. The notion of universe of content and the notion of scalability are quite distinct. If a universe of content is not scalable, it can sometimes be broken down into sub-universes, some of which may be scalable separately. If a universe is not scalable for a given population of people, then it is not meaningful to assign a single rank order to the people with respect to the total content. Indeed, if arbitrary scores were assigned to non-scalable data, intensity analysis should find that there was no U- or J-shaped intensity function and no invariant zero-point for dividing the population into positives and negatives.

An example of such a non-scalable case is the one next to be given. It will also illustrate the two-part intensity technique. The content for this second example concerns another textbook used in the same course as the first. The 50 students in the class were asked the following questions about *Black Metropolis* (by Drake and Cayton):

#### *Black Metropolis*

##### Questions

1. (a) On the whole, as textbooks go, how good do you think *Black Metropolis* is? (Check one answer)
 

Very good	Good	Fairly good
_____ 5 _____	_____ 4 _____	_____ 3 _____
Passable	Not very good	Terrible
_____ 2 _____	_____ 1 _____	_____ 0 _____
- (b) How strongly do you feel about this? (Check one answer)
 

Very strongly	Pretty strongly
_____ 3 _____	_____ 3 _____
Somewhat strongly	Not strongly at all
_____ 1 _____	_____ 0 _____
2. (a) In your opinion, does *Black Metropolis* present a good sociological analysis of the Negro community in Chicago?
 

An excellent analysis
_____ 5 _____

- \_\_\_\_\_ A very good analysis 4  
 \_\_\_\_\_ A pretty good analysis 3  
 \_\_\_\_\_ It has only a few good points 2  
 \_\_\_\_\_ Not a very good analysis 1  
 \_\_\_\_\_ A pretty bad analysis 0
- (b) How strongly do you feel about this?
- \_\_\_\_\_ Very strongly 3 \_\_\_\_\_ Pretty strongly 2  
 \_\_\_\_\_ Somewhat strongly 1 \_\_\_\_\_ Not strongly at all 0
3. (a) To what extent does the book afford the student a real insight into the problems of race relations in Chicago?
- \_\_\_\_\_ Not much at all 0  
 \_\_\_\_\_ A somewhat limited insight 1  
 \_\_\_\_\_ Fairly good insight 2 \_\_\_\_\_ A good insight 3  
 \_\_\_\_\_ A very good insight 4 \_\_\_\_\_ An excellent insight 5
- (b) How strongly do you feel about this?
- \_\_\_\_\_ Very strongly 3 \_\_\_\_\_ Pretty strongly 2  
 \_\_\_\_\_ Somewhat strongly 1 \_\_\_\_\_ Not strongly at all 0
4. (a) In general, how well does the book organize and present its material?
- \_\_\_\_\_ Very poorly 0 \_\_\_\_\_ Not very well 1  
 \_\_\_\_\_ Fairly well 2 \_\_\_\_\_ Quite well 3 \_\_\_\_\_ Very well 4
- (b) How strongly do you feel about this?
- \_\_\_\_\_ Very strongly 3 \_\_\_\_\_ Pretty strongly 2  
 \_\_\_\_\_ Somewhat strongly 1 \_\_\_\_\_ Not strongly at all 0
5. (a) Some parts of *Black Metropolis* emphasize statistical data and other parts quote personal interviews a great

deal. Do you believe that the authors have succeeded in blending these data properly, or have they failed?

Succeeded very well 4  
 \_\_\_\_\_  
 Succeeded pretty well 3  
 \_\_\_\_\_  
 Succeeded at least more than they have failed 2  
 \_\_\_\_\_  
 Pretty much failed 1 \_\_\_\_\_ Definitely failed 0

(b) How strongly do you feel about this?

Very strongly 3 \_\_\_\_\_ Pretty strongly 2  
 \_\_\_\_\_  
 Somewhat strongly 1 \_\_\_\_\_ Not strongly at all 0

6. (a) Some students complain that the textbook often makes fuzzy statements, so that it is not clear what position it takes or what it is driving at. To what extent do you agree with this complaint?

Completely agree 0  
 \_\_\_\_\_  
 Agree for the most part 1 \_\_\_\_\_ Undecided 2  
 \_\_\_\_\_  
 Disagree 3 \_\_\_\_\_ Completely disagree 4

(b) How strongly do you feel about this?

Very strongly 3 \_\_\_\_\_ Pretty strongly 2  
 \_\_\_\_\_  
 Somewhat strongly 1 \_\_\_\_\_ Not strongly at all 0

7. (a) Do you think *Black Metropolis* is good enough to be kept as a textbook for this course?

Definitely yes 4 \_\_\_\_\_ Yes 3 \_\_\_\_\_ Undecided 2  
 \_\_\_\_\_  
 No 1 \_\_\_\_\_ Definitely not 0

(b) How strongly do you feel about this?

Very strongly 3 \_\_\_\_\_ Pretty strongly 2  
 \_\_\_\_\_  
 Somewhat strongly 1 \_\_\_\_\_ Not strongly at all 0

Each question is in two parts. The first part is to study content, and the second part is to study intensity. Notice that

the number of categories in the content parts are not uniform from question to question. It is not essential for a scalogram analysis that there be any uniform format for the questions. In the same series of items, some can be trichotomies, some can have six categories, and some can have two categories, etc. Nor is the wording of the categories of special importance. Short phrases or long phrases, etc., can be used. Five and six categories were used in the present example because it was suspected in advance that the students would give apparently favorable answers to all questions put to them, so the apparently favorable responses were made more numerous in the check list of answers in order to help obtain differentiation in rankings.

The Cornell technique was used to analyze the content parts of the seven questions on *Black Metropolis*. The first trial weights are those indicated with the questions, and the first trial scalogram is shown in Table 6. All of the items were found to have so much error that they required dichotomization. The combinations of categories used and the results of the second trial are shown in Table 7. There is still too much error in Table 7. Several of the questions have more error than non-error. We therefore judge the total content not to be scalable, since no further trials can be made when all items are dichotomous.

Therefore, we cannot speak of degrees of "favorableness" of opinion about *Black Metropolis* for this class of students. We cannot say that one student likes the book better than another student. He may like it better in one of the aspects and not better in another. There is apparently not a single ranking possible in the total content studied by the questionnaire. If the study were to be carried further, what would be done would be to try to break the content down into sub-areas, make up a dozen or so questions for each of the sub-areas, administer the sub-areas to the class, and analyze each separately by scalogram analysis. Such a further study was not made for this present example.

*The Two-Part Intensity Technique.*—Since the total content is not scalable, it is hardly worthwhile to study intensity. However, in order to see how the two-part technique operates,

TABLE 6

**Score**

TABLE 6—(Continued)

Score	1					2					3					4					5					6					7																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5



TABLE 7  
BLACK METROPOLIS  
Second Trial, Content

[illegible]



TABLE 8—(Continued)

Score	1			2			3			4			5			6			7									
	3	2	1	0	3	2	1	0	3	2	1	0	3	2	1	0	3	2	1	0								
14	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x								
14	x	x	x																									
14																												
13																												
13																												
13																												
13																												
13																												
13																												
12																												
12																												
12																												
11																												
11																												
11																												
11																												
11																												
11																												
10																												
10																												
10																												
9																												
3																												
0																												
Freq.	9	31	8	2	20	23	3	4	19	23	7	1	13	25	10	2	15	25	9	1	10	18	16	6	17	21	8	4

let us go through with it anyhow. Each part (b) of the 7 questions on *Black Metropolis* was weighted according to the weights indicated in the list of questions above, and trial intensity scores were obtained thereby. Intensity again seems to be a quasi-scale. Obtaining a quasi-scale, however, has no bearing on the scalability of the content. The scalogram for the trial intensity is shown in Table 8. Plotting the trial intensity scores against the second trial content scores yields the scattergram in Table 9. Again, the frequencies in boldface in

TABLE 9  
*BLACK METROPOLIS*  
*Scattergram of Intensity and Content*

Intensity	Content (Second Trial)							Total
	0	1	2	3	4	5	6	7
21							1	1
20							1	2
19						1		1
18					1		1	3
17					2	1	1	5
16							1	4
15				1	2		1	6
14	1			1		1	1	5
13				1		2	1	6
12		1	1				1	3
11				1	1	1	2	6
10				1			2	3
9	1							1
3							1	1
0								1
Total	2	1	1	5	6	6	14	50

each column indicate the median position for intensity for the respective columns.

As stated previously, fifty cases are far from sufficient to obtain stable column medians when there is a substantial intensity error present, which seems to be the case here. However, we have strong reason to believe that the absence of a U- or J-shaped curve of medians in Table 9 is not merely due to sampling error, but rather to the fundamental lack of scalability of the content.

*An Intensity Curve from a Final Survey.*—To give the reader a picture of what final results will look like in practice in

a complete study, we present some data from a study by the Research Branch. Ten questions were asked of a cross-section of 1800 enlisted men with respect to the expression of job satisfaction in the Army. The content was found to be scalable. Intensity was obtained by the two-part technique. The relationship between intensity and content is shown in Table 10.

TABLE 10  
*An Example of the Intensity Function:  
Job Satisfaction in the Army*

Intensity Score	Content Score											Total
	0	1	2	3	4	5	6	7	8	9	10	
8	23	46	27	33	22	19	24	42	25	23	24	308
7	7	24	31	26	33	31	22	40	21	15	5	255
6	1	7	29	17	30	24	35	42	15	11	.	211
5	6	14	14	29	20	34	27	34	19	10	..	197
4	2	3	15	17	32	33	36	36	10	1	..	185
3	..	1	17	19	22	29	33	25	11	1	..	158
2	1	4	9	19	25	34	31	32	1	4	1	161
1	..	2	2	12	35	39	38	30	8	.	1	167
0	..	3	7	12	29	43	33	26	3	1	..	157
Total	40	104	151	184	248	286	279	307	103	66	31	1,799

The frequencies in boldface in the columns show the median intensity for the respective columns. Content score 5 seems to be approximately the zero interval. Men to the right of this score can be said to have *positive* job satisfaction, and men to the left to have *negative* job satisfaction.

In conclusion, it should be pointed out that the intensity curve provides not only an objective zero-point, but also a picture of the relative strength with which an attitude or opinion is held. Differing shaped curves, when plotted on the percentile metric, show differing degrees of sharpness in the division of attitudes or opinions. Illustrations of this will be given in future publications.<sup>8</sup>

<sup>8</sup> See Guttman, Louis and Suchinyl. "Intensity and a Zero Point for Attitude Analysis," *American Sociological Review*, XII (1947), 57-67.



# AN EMPIRICAL STUDY OF THE RELATIONSHIP BETWEEN ITEM VALIDITY AND INTERNAL CONSISTENCY<sup>1</sup>

WILLIAM A. OWENS

Iowa State College

## *Introduction*

TECHNIQUES for evaluating tests and the items which compose tests found increased application throughout the war period just passed. The opportunity to use certain of these techniques first hand gave valuable experience to many who would not otherwise have had it. Conversely, however, there was seen an occasional tendency to uncritically adopt a given statistical procedure and then to apply it beyond its proper limitations. This present article is written as a criticism of one such practice.

Various techniques of item analysis are basic in psychometric work today. Of these none seems more commonly recognized and used than the method of internal consistency. It is natural that psychologists should have had frequent recourse to this method for—as in the field of personality measurement—it is often difficult or even impossible to find a satisfactory external criterion for use in the selection of items.

It is, nevertheless, somewhat disturbing to hear a query as to the validity of a measuring instrument answered with a statement to the effect that the mean internal consistency of the items is, say, 0.40. This tendency to identify internal consistency with validity and to speak of the two as roughly equivalent is an understandable, but all too frequently encountered error. It is this assumption of near identity which it is presently proposed to evaluate.

---

<sup>1</sup> The author gratefully acknowledges the suggestions and criticisms of the Doctors Charles I. Mosier and Martin F. Fritz.

For the purposes of this discussion item validity ( $I_v$ ) may be defined in terms of the biserial correlation of an item on a test with some external criterion of what the test purports to measure. (In the present case school grades were the criterion.) The index of internal consistency ( $I_c$ ) may be given a comparable definition in terms of the biserial correlation of an item in a test with the total score on that test. Internal consistency ( $I_c$ ) carries this meaning throughout the present article.

It must seem immediately obvious from this definition that  $I_v$  and  $I_c$  cannot be identical unless the external criterion for the test and total score on the test are correlated 1.00—a situation often dreamed of and never encountered. A question may then be raised as to the nature of the hypothesis which would allow their being viewed as even roughly equivalent.

The line of reasoning involved runs somewhat as follows:

Let it be assumed that the total score on a test shows a correlation with some appropriate external criterion. It may then be argued that: (a) the majority of items in the test measure what it was intended they should measure, or are valid, and (b) if the test were analyzed for  $I_c$ , and thus "purified," a large percentage of its items would function like those already operating with some validity and the validity of the total test would be correspondingly increased.

In order to evaluate this hypothesis it is necessary to state precisely what the effect is upon a test of selecting items for high internal consistency. As Richardson (5) has stated, "In a test of uniform difficulty, the correlation of an item with the test is proportional to the *average* correlation of that item with *each item* of the test." It then follows that "*The rejection of items whose correlations with the test are relatively low raises the average intercorrelations of the remaining items.*" This result may be qualitatively described as purifying, narrowing, intensifying, unifying, or homogenizing the test. As Mosier (4) has so clearly pointed out, this will only surely tend to increase validity if "*the original list of test items measures a single trait*"—a situation which is rarely obtained in practice. If, however, several traits are involved and validity is estimated from a complex criterion, then "narrowing up" the test may "shrink it away" from the criterion.



For example, let the uncorrelated traits A and B be imagined with each one contributing one-half of the criterion variance. Let it be further supposed that an experimental form of a test to measure these two traits is composed of 300 items, 200 of which are measuring trait A and 100 of which are measuring trait B. Then, if the final form of the test is to be composed of 100 items, and if the 100 are selected by the method of internal consistency, there will be a tendency for this two-to-one proportion to shift even further toward the inclusion of type A items and the exclusion of type B items. This follows from the fact that the type A items will have more "same-type" items in the total score with which each item is correlated, and will thus tend to have higher *mean* item intercorrelations ( $I_o$ 's). This would obviously reduce validity in proportion to the extent to which it resulted in the prediction of only one-half of the criterion variance—the trait A half. A multiple-trait test to predict a simple criterion could incur comparable damage if, in the selective process, an increasing number of items were pulled to represent a trait *not* represented in the criterion. However, regardless of purity or complexity of criterion, no such situation could arise in the selection of items if the original list covered only a *single-trait*.

That the effects hypothesized *do* occur in practice will be borne out by the experience of those who have constructed tests. It is not unheard of for a test to have a lower validity coefficient following an  $I_o$  analysis than it did before the analysis—this in the absence of other factors to explain the change. The writer recently had such an experience in the development of a General Physics Test designed to predict success in a naval Pre-Radar School. There the test versus grade-point correlation coefficient dropped 8 points following a revision in which items were selected for high  $I_o$ . Reliability did not suffer, and the observed difference was statistically significant. Apparently a narrower coverage of the subject matters of physics was responsible.

It thus appears, both theoretically and practically, that the validity of a multiple-trait test may suffer if the items in the test are selected only for  $I_o$ . It is suggested that a causal fac-

tor is the "narrowing" effect of such selection which can result in more and more reliable prediction of fewer and fewer segments of the criterion.

A further illustration or two may be warranted by way of clarification. For example, consider this *reductio ad absurdum*. Let us imagine a test composed of 100 identical items, or of 100 items operating in identical fashion and designed to predict a complex criterion. Here the item versus test correlations would be 1.00 and  $I_r$  would be at a maximum. However, the validity of such a test would be the validity of a single item, and such limited coverage would surely result in dubious validity—hardly a maximum. A point of importance finds illustration here; it is that an item analysis for  $I_r$  actually contributes more to the *reliability* of a test than to its *validity*.

From another aspect, it has long been an axiom in test construction that the tests of a battery should correlate *low* with each other and *high* with the external criterion of validity. The items of a multiple-trait test would seem to bear the same relationship to the test as the tests of a battery do to the battery (assuming a single, additive score on the battery). If this is a fair analogy, then the items of a test should *not* be chosen for *high* mean intercorrelations ( $I_r$ ), tending towards duplication and overlap, but for *medium* to *low* intercorrelations and high criterion validity. In the absence of a suitable external criterion, the not uncommon practice of selecting items for high  $I_a$  *alone* would accordingly appear to have little to recommend it.

In this connection it is germane to refer to Horst (1) who describes an item selection technique for maximizing validity. He sets up a hypothetical situation in which the first 50 items of a 180 item test to be drawn from a pool of 500 items would be those which correlated positively with the criterion and *negatively with the total test*.

In dilating upon the relative independence of  $I_a$  and  $I_r$ , it is impossible to improve on Richardson (5). He says analysis for  $I_a$  . . . "will tend to make the test more pure or homogeneous, in the sense of conserving those items which have the largest intercorrelations. This is the only sense in which it

may be said that the conserved items are more 'valid' than the rejected items." Lorge (3) summarized well when he pointed out that lack of  $I_c$  may *invalidate* a test or questionnaire, but that the presence of  $I_c$  will not necessarily validate it.

The writer was interested in supplementing logic on this issue with empirical evidence, and gathered the data which follow with this end in view. It is unlikely that the tests employed are factorially pure, but they are probably typical of much data to which the method of  $I_c$  is applied. The magnitude of the obtained mean relationship between  $I_c$  and  $I_v$  surely suggests that these are relatively independent criteria for the evaluation of a measuring instrument.

#### *Present Evidence*

*Problem.*—To adduce some evidence as to the relationship between the indices of internal consistency ( $I_c$ ) and those of item validity ( $I_v$ ).

*Method.*—Two treatments were accorded the same data. The first was to run product-moment correlations, by tests, between the paired indices of item validity and those of internal consistency. The second was to throw the 204 items constituting the tests into a four-fold table which would show the frequency of agreement or disagreement between the two series of indices, with each series split into two categories at its mean.

All of the biserial correlations obtained as indices of  $I_v$  or  $I_c$  were computed from data on a systematic or proportionate sample of 370 cases drawn from a population of approximately 1,000 Navy enlisted men. The four tests concerned were taken from the Navy Basic Test Battery. (Table 1) They had reliabilities of 0.81 to 0.95 and criterion correlations of 0.37 to 0.49.

Grades at the Navy Basic Engineering School were used as the criterion of item validity. The odd-even weeks reliability of these grades ranged between 0.80 and 0.90 over a period of nearly six months. Such coefficients are large enough to indicate that unreliability of the criterion is *not* a major factor attenuating the correlations which represent item validity.

It was, however, necessary to obtain corrected estimates of

TABLE 1  
*Correlation of Item Validity and Internal Consistency*

Test	Rel't'y	$r$ with criterion	No of items	$r$ of $I_v$ & $I_c$ (uncorrected)	Significance
1. G.C.T.*	0.98	0.38	100	0.42	$P < .01$
2. Read'g.	0.89	0.37	30	0.50	$P < .01$
3. Arith.	0.88	0.49	30	0.25	$P \geq .05$
4. Mech'l. Comp'n.	0.81	0.46	44	0.36	$P < .02$
				Av. 0.39	

\* G.C.T. is a General Classification Test composed of completion (30), opposites (30), and analogies (40) items

the magnitudes of the four correlations between the indices of  $I_v$  and  $I_c$ , since all of the items for which validities *later* became available were *originally* selected for internal consistency. The resulting restriction in the range of these latter values had the effect of lowering their observed correlation with item validity. Since indices of internal consistency were available for all items in both the experimental and final forms of the several tests, it was possible in each case to make an accurate determination of the amount of restriction involved.

Kelly's (2) formula was then applied to obtain an estimate of the true magnitude of the  $I_v$  versus  $I_c$  relationship existing within each test. This formula corrects the correlation coefficient for the effect of a measurable restriction in one variable.

*Results.*—The results of this study are incorporated in three tables. Table 1 deals primarily with the uncorrected correlations between item criterion validity ( $I_v$ ) and internal consistency ( $I_c$ ). It should be noted that these coefficients range between 0.25 and 0.50; their weighted average is 0.39.

Table 2 supplements Table 1 and shows a corrected value for

TABLE 2  
*Correction for Restriction in Range of  $I_c$  Values*

Test	$r$ of $I_v$ & $I_c$ (original)	Initial $\sigma$ of $I_c$	Final $\sigma$ of $I_c$	$r$ of $I_v$ & $I_c$ (corrected)	Significance
1. G.C.T.	0.42	17.45	9.62	0.64	$P < .01$
2. Read'g.	0.50	12.21	8.94	0.62	$P < .01$
3. Arith.	0.25	16.06	10.93	0.36	$P \geq .05$
4. Mech'l. Comp'n.	0.36	17.94	14.70	0.42	$P < .01$
				Av. 0.55	

each of the original correlations; it also shows the initial and final standard deviations which evidence the restriction in  $I_o$  values for which the correction is made. It may be observed that these coefficients have increased considerably in magnitude and that their weighted average is now 0.55. However, the predictive efficiency of a correlation of 0.55 is not great. In predicting item validity from internal consistency with this coefficient, one might expect to do  $16\frac{1}{2}$  per cent better than chance. Such a prospect would hardly be encouraging if one intended to determine the former variable through measurement of the latter.

Table 3 shows the extent of agreement between the item

TABLE 3  
*Overlap and Discreteness in the  $I_v$  and  $I_o$  Criteria*

Below mean $I_o$ Above mean $I_v$	32	Above mean $I_o$ Above mean $I_v$ 66
Below mean $I_v$ Below mean $I_o$	70	36 Below mean $I_v$ Above mean $I_o$

criteria of validity and internal consistency. It will be observed that the non-predicted cells contain one-third of the items ( $32 + 36$ ). It will also be noted that if only items above the mean internal consistency were chosen for inclusion in a test, there would still be over one-third of these items below the mean validity. Such a result could hardly be interpreted as backing the impression that criterion validity and internal consistency are near equivalents, although it does point again to the fact that they are significantly correlated.

It is recognized that these data are not adequate to warrant any sweeping generalizations. Different tests administered to different populations would yield different estimates of the magnitude of the relationship under consideration. However, it does seem to be indicated that in an adequate Navy population, with a fairly representative sampling of tests and with a reliable criterion, the biserial correlation of an item with total score on a test may *not* safely be assumed to be equivalent to

the correlation of the item with an external criterion of the thing to be measured.

### *Conclusions*

In this population and with these tests, the following conclusions may be drawn:

1. The average estimated correlation between the indices of item validity and those of internal consistency was 0.55.

2. With the four-fold table of item categories employed, there was 33 1/3 per cent *disagreement* between the two indices named above.

### REFERENCES

1. Horst, P. "Item Selection by Means of a Maximizing Function." *Psychometrika*, I (1936), 229-244.
2. Kelly, T. L. *Statistical Method*. New York: Macmillan Company, 1923. p. 225.
3. Lorge, I. "Personality Traits by Fiat: II. A Correction." *Journal of Educational Psychology*, XXVI (1935), 652-654.
4. Mosier, C. I. "A Note on Item Analysis and the Criterion of Internal Consistency." *Psychometrika*, I (1936), 275-282.
5. Richardson, M. W. "Notes on the Rationale of Item Analysis." *Psychometrika*, I (1936), 69-75.

# A NEW APPROACH TO ORAL TESTING

WILLIAM BRODY  
New York City Department of Health

and

NORMAN J. POWELL  
College of the City of New York

AMERICAN Civil Service administration has exhibited curiously uncoordinated attitudes towards the oral test. On the one hand, personnel administrators have insisted upon the importance of measurement in high-level positions of such attributes as personality and the ability to get along with people; the oral test, it is argued, is the sole available device for appraising these personal qualifications. On the other hand, an imposing experimental literature professes to demonstrate that the oral test is an unreliable tool suggestive of flipping slightly twisted coins to decide the rank order of candidates.

Thus practice ignores research. Research has produced an extensive literature and, in practice, the oral examination goes its adventitious way wherein candidates who would have been marked one way by the Examining Panel Alpha are rated in another way by the Examining Panel Beta.

Essentially, the oral test problem (identical with any other in Civil Service measurement) is to evaluate, reliably and validly, desired personal attributes while maintaining public confidence in the adequacy of the selection procedure.

This paper describes a new kind of oral test recently administered by the New York City Department of Health. The technique used represents a sharp break with convention. It is based on methods which were developed by the British Army in selecting officers during World War II and adopted by the British Civil Service Commission for selecting administrative and foreign service personnel, and which were used on tryouts

conducted under the direction of Milton M. Mandell of the United States Civil Service Commission.<sup>1</sup>

Perhaps others may wish to repeat, with such modifications as may seem appropriate to differing conditions, the group oral performance test conducted in New York City. Pertinent details follow.

*The Position.*—Health-officers-in-training are nominated by the Department of Health, subject to the approval of the Municipal Civil Service Commission, for an eight-month course in a school of public health, leading to the degree of Master of Public Health. The remainder of the two-year tenure of this position is devoted to an orientation course within the Department of Health. Tuition fees and a stipend are paid for out of funds which are made available through the United States Public Health Service.

Those who complete the program are eligible to participate in a competitive Civil Service examination for the position of Assistant District Health Officer, a position involving second-line responsibility for the comprehensive program of a health district for a population of approximately 250,000.

The minimum requirements for those wishing to become health-officers-in-training are: United States citizenship, eligibility for a license to practice medicine in New York State, and one year of internship or residency in an approved hospital.

*The Candidates.*—Candidates for these positions are secured by means of a continuous recruiting program. Each candidate is interviewed whenever he presents himself. This informal interview lasts for approximately an hour and serves the dual purpose of discouraging those who are obviously not qualified and of encouraging those who seem qualified.

*The Preliminary Examination.*—All candidates who were sufficiently interested to file their applications after this preliminary interview (during the four-month period after the last test was given for this position) were invited to an "experience interview test." Each of the twelve applicants who responded was interviewed for a period of approximately 15 minutes by a

<sup>1</sup> Mandell, Milton M. "The Group Oral Performance Test." *Public Personnel Review*, VII (1946), 209-212.



four-man examining panel. On the basis of this examination four candidates who were obviously not qualified for the position were eliminated.

*The Preparatory Material.*—Each participant was given a set of instructions which pointed out that a “group oral performance test” would be held five days later, and that the candidates would be responsible for (1) preparing group recommendations to the Commissioner of Health on a specific medical problem (the control of rabies; detailed information was supplied) and (2) delivering a short talk on one of eight carefully delineated subjects, with each candidate making his own choice from among the eight.

*The Group Oral Performance Test.*—The rating board for this test consisted of the same four examiners who participated in the experience interview. Two of them sat along each of the long walls of the examination room.

As the candidates arrived they were allowed to select their own seats around a conference table. As soon as all eight<sup>2</sup> were seated, a sheet of paper was distributed to each. This contained the same information concerning the rabies control problem as they had previously received, with the addition of the following two paragraphs:

You will be rated on the value of your contribution to the recommendations made, your approach to the problem, your group participation, and your manner and speech in presenting your opinions.

**SPEAK CLEARLY SO THAT EVERY ONE CAN HEAR YOU.** The group itself will decide how the discussion is to be carried on. Start the discussion as soon as every one has finished reading this statement. Continue until a signal is given to stop.

No additional directions were given and no questions were asked by the examiners.

Within two minutes one of the candidates took the lead and before long all of them were participating in the discussion. After about thirty minutes they agreed to select one of their members as secretary.

---

<sup>2</sup> If more than ten or twelve candidates had been involved it would have been necessary to divide them into subgroups.

About thirty minutes later the examiner in charge announced the conclusion of this part of the test and a five-minute recess was taken.

The examiners then reversed their seating arrangement and the candidates decided to keep the same seats they had had before. The material which was distributed to the participants at this point consisted of the same explanatory information concerning the eight optional subjects as had been distributed at the first session, with the following added comments:

After you have finished, the members of the group will assume that they constitute the appropriate audience for your remarks and will ask pertinent questions.

The order in which each individual is to speak will be determined by the group. The group will also decide how much time should be allowed for general discussion after each talk.

Start the discussions as soon as every one has finished reading this statement.

**SPEAK CLEARLY.** You will be rated both on your own presentation and on the questions you ask.

After a very brief discussion, the group selected the first speaker and determined that the order of seating would constitute the order of speaking.

When the initial speaker had concluded his talk, the group was at a loss as to how to proceed. It was obvious to the examiners that no candidate had read the final four paragraphs of the explanatory material, and it became necessary to direct them to do so. If this development had been foreseen, it would have been a simple matter to place this important information at the beginning of the instruction sheet.

Thereafter the program continued without complications. It may be of interest to note that every candidate exceeded the length of time prescribed in the preliminary information (3-5 minutes); in one case the elapsed time was 15 minutes. Yet the group held rigidly to its self-imposed limit of 5 minutes for general discussion after each talk.

The examiners had prepared an additional problem for group discussion, but they unanimously agreed that they had already learned enough about the applicants. The examination was concluded slightly less than three hours after it had begun.

It is to be noted that prior to the test the candidates had been informed of the problems to be considered. This step was taken to eliminate the advantage that might otherwise accrue to some candidate or candidates because of fortuitous special knowledge of the subject.

The test was not intended to measure information. In the specific situation, a common basic knowledge of medical theory could be assumed because of the minimum requirements; no further knowledge was considered important for selection purposes because successful candidates would undergo a two-year training course. For positions where subject matter is an important consideration, a preliminary written test would presumably be indicated.

The ability of the candidates to gather appropriate data in the five-day period between tests might also be considered a significant element in rating. On the other hand, inequality among candidates might actually be a measure, not of ability to collect facts, but of time available for this activity.

The candidates were not given specific information in advance about the nature of the test. For example, they were not informed that there would be a general discussion after each individual talk.

An interesting detail in the test administration was the use of name cards on the table. Each candidate was furnished with a folded sheet of heavy cardboard with his own name in large letters on each side of a center fold. This had the double advantage of enabling the group to address each other by name, thus minimizing the artificiality of the situation, and of protecting the examiners against assigning ratings to the wrong people.

To prevent any candidate from receiving the psychological advantage of being seated at the head of the table from the beginning of the proceedings, chairs were placed only at the sides of the table. If the group had selected a chairman, he could, of course, have moved around to the head of the table.

The examiners were seated far enough away from the table to allow the participants to forget (at least occasionally) that they were present, but not too far away to miss anything. To

avoid any possible bias created by looking at the backs of some candidates, they reversed their positions in the middle of the test.

The degree to which each candidate participated was an important element of the examination. The technique of a brief individual talk was incorporated in the test specifically to insure adequate consideration of even those candidates who might not participate in the general group discussion. As it turned out, there was one such candidate.

Each of the examiners rated the candidates on six factors. Each examiner determined his final rating of each candidate by adding the six factor ratings and the experience rating (arrived at in the preliminary examination) and then dividing by 7. The six factors and the elements considered under each factor were:

1. *Appearance and Manner.* Poise, physical alertness, nervousness, attentiveness, mannerisms.
2. *Speech.* Power of expression, vocabulary, diction, modulation.
3. *Attitude towards Group.* Tact, cooperation, ability to mix, flexibility.
4. *Leadership.* Ability to assume lead without giving offense, acceptance by group.
5. *Contribution to Group Performance.* Team-worker or prima donna, awareness of objectives of group discussion, ability to reconcile differences.
6. *Scientific Approach.* Ability to marshal data, awareness of implications, ability to reason, ingenuity, mental alertness, judgment.

Exact quantitative data are not available for the evaluation of the group performance test.<sup>3</sup> However, it is possible to set down a number of apparent advantages. The rationale in opposition to each point made is recorded parenthetically.

<sup>3</sup> It was the unanimous opinion of the four examiners, as well as the independent judgment of each of three observers with testing experience but no medical background, that three of the candidates should be considered as not qualified for the training course, while the remaining five showed considerable promise. There was less agreement concerning the rank order of the five who were considered qualified. With the exception of one examiner, however, all rated two specific candidates in either first or second place.

1. The group performance test enables the rating examiners to observe each candidate in action for a period of three hours. In the same amount of examiner time each applicant could be granted an individual interview of only twenty minutes.

(But raters did not observe each candidate for three hours. Rather, three hours of observing time was distributed among the candidates, necessarily in rough relation to such items as the time taken for speaking by each individual, interesting physical or behavioral characteristics of candidates, visual and auditory considerations, and like attention-arresting factors.)

2. It permits each examiner to devote full time to observing, listening, and taking notes.

(Examiners' attention must lag intermittently and unpredictably. Without the stimulation of continuing verbal contact between candidate and rater, the mind of the examiner may wander increasingly as the test goes on.)

3. It eliminates any tendency on the part of the examiners to use the oral interview as a means of impressing the other panel members with their own knowledge and skill in questioning and subject-matter.

(Any such tendency posits a type of examiner best dealt with by eliminating him from the examining process, or at least by appropriate preliminary briefing. Deprived of the opportunity to strut, such an examiner will, if allowed to continue, merely find other undesirable outlets for his peculiarities.)

4. It prevents any loss of reliability caused by the use of different questions for different candidates as well as by the information given to later candidates by those examined earlier.

(This assumes that the candidates are so few in number as to be expeditiously handled in one group session. Furthermore, in the ordinary oral test, it is quite possible to arrange to use the same questions, if desired, for different candidates. In the group oral performance test itself, standardization of detail tends to be at a minimum, inasmuch as the candidates may roam almost as they will.)

5. It minimizes the effect of the inevitable lack of continuous concentration on the part of examiners.

(But it tends to maximize the probability that the ex-

aminers' concentration will not be continuous. The fact that the examiner is, in a sense, himself under observation in the usual oral test keeps him, at least seemingly, wide awake.)

6. It provides a more natural situation than the usual question-and-answer contact between a candidate and his examiners.

(It is certainly unnatural for adults to be required to discuss a particular matter under the silent and somewhat remote inspection of examiners. The usual question-and-answer contact possesses the virtue of being usual; candidates are accustomed to such a situation and are more likely to regard it as natural.)

7. It eliminates the suspicion on the part of any candidate that other candidates may be received more kindly and may be given easier questions. It may even convince him that some other candidates are better qualified than he is.

(This is arrant assertion, not evidence. Suspicious candidates could in any case imagine partisanship or other bias on the part of the examiners.)

8. It eliminates the following dangers in the conventional oral interview situation:<sup>4</sup>

a) "There exists a tendency for panel members to slant their attention to the candidate's response to their own questions."

b) "It has been found that the rating given to any candidate tends to carry over positively to the following candidate. The consequence is a kind of halo effect."

(a) In the conventional interview we can be quite sure that the rater listens to the answers to his own questions, if nothing else. A similar near-guarantee is absent in the group situation.

b) Perhaps kindred biases occur in the group test. Where individual candidates sit is determined by chance, but where they sit may be correlated with their rating.)

9. It provides very valuable information concerning the attitude of each candidate toward the other members of the

<sup>4</sup> Pointed out by Norman J. Powell and Harold Levine in a mimeographed *Manual of Procedures in Oral Tests* prepared for the Civil Service Commission of the City of New York.

group, as well as of his reaction to their attitudes. This is particularly important in testing for positions where group discussions and conferences are essential.

(But the situation may be regarded as so unreal as to yield perverted findings. In the real conference situation, each member generally has a defined status. Very importantly, conferees probably have prior acquaintance with one another and the good conferee will use to advantage the information which he possesses about the personalities of his colleagues in reaching a predefined objective.)

10. It presents specific evidence concerning the ability of each candidate to be a leader in a group.

(More accurately, the group test provides specific evidence of the candidate's ability to lead the particular group in whose deliberations he participates on a specified subject under the conditions set in the test. It remains possible that varying elements of the situation may result in varying the performance of candidates.)

11. Those who participate (examiners as well as candidates) find it more interesting than the individual interview.

(Interest is not to be confused with validity and public relations values. The latter are the significant aspects of a test.)

12. It requires no skill in asking questions on the part of examiners.

(Here, in fact, is one of the greatest weaknesses of the group examination. Opportunity is denied to the examiner to explore the candidates' statements, to follow up leads in order to verify the apparent competence or incompetence of the applicant.)

Obviously, there is much to be said for and against the group oral performance test. The test is promising in many ways; it has decided weaknesses. Perhaps the best argument which can be advanced at the present time in support of the group test is a negative one—it is unlikely to be as bad as the orthodox oral interview test.

Certainly additional experimentation by public personnel agencies with the group examination is most desirable. To be sure, this is a feeble conclusion and bears resemblance to the

traditional position of the liberal whose two feet are firmly planted in mid-air. Yet the fact appears to be that all that can confidently be said about the group oral performance test is that its possible utility is worth exploring.



# THE VALUES OF AN EDUCATIONAL GUIDANCE CLINIC<sup>1</sup>

JOSEPH C. HESTON

DePauw University

IN July, 1938, Allegheny College inaugurated a new venture in educational practice that has not yet received the degree of attention and consideration it richly deserves. An Educational Guidance Clinic was established to assist pre-college students in their efforts at reaching certain common educational and vocational decisions. This clinic has been well described in the article by Anderson (1). It is the purpose of this brief report to focus attention on this type of educational venture by seeking to point out the theoretical values that may be gained from guidance clinic procedure. The results obtained from an opinion survey conducted among students participating in a similar clinic program at DePauw University in June, 1946, will also be presented. We hope to stimulate interest in this project, so that additional colleges will be encouraged to offer comparable clinic programs.

## *The Educational Guidance Clinic at DePauw University*

It is desirable to provide a brief orientation into the nature of the Educational Guidance Clinic as it has operated on the campus of DePauw University. The clientele to whom the services of the clinic have been offered are the graduating high-school seniors in schools from the area generally contributing to the DePauw enrollment. A brief mimeographed two-page circular is sent to each student who makes an inquiry concerning entrance to the DePauw admissions office. Copies of the cir-

---

<sup>1</sup> The writer is personally indebted to Dr. Hurst R. Anderson, now President of Centenary Junior College, and to Dr. Guy E. Buckingham, Head of the Department of Psychology and Education at Allegheny College, for interviews providing descriptions of the founding and operation of the Allegheny Educational Guidance Clinic.

culars are also supplied to principals and to counselors of the various high schools located within our area. Students are advised that this is an *impartial* service open to all high-school graduates, regardless of whether they have chosen to enter DePauw or some other college. It offers a personalized guidance situation in which the student can plan, well in advance of actual registration, a suitable course of college study. This early introduction to college procedures and problems is in itself a valuable experience, in addition to the careful individual counseling received by each student.

The participating students are given two days of standardized tests covering scholastic aptitude, reading skills, achievement in the four essential areas of high-school work, vocational interest, and personality adjustment. Complete physical examinations are also provided by the University Health Service. All assembled data, including pertinent autobiographical information from each student, are first given staff consideration by a small group of well-trained counselors. Conclusions reached at this staff conference are then interpreted individually by one of the counselors to the student and to his parents. This final case conference is the concluding one of a series of counselor-student contacts and comes on the fourth day of the clinic. Although this concluding conference is the only *formal* interview, the student and his counselor have had several meals together and have spent an evening together at a planned social party in the gymnasium recreational rooms. Further, the student-parent-counselor combinations are first brought together at an informal picnic involving all clinic participants. Much significant "interviewing" is carried on unobtrusively in these contacts, under conditions frequently more productive of clinical evaluation than customarily obtains in routine student counseling.

The participating counselees are housed and provided meals by the regular campus facilities. As indicated above, various recreational opportunities and group activities are provided to supplement the testing and counseling, and afford further opportunities for *informal* contacts. A fee of twenty-five dollars (\$25.00) is charged each student. This fee covers all of the

expenses of the clinic and is designed to just meet the actual cost of the various services rendered. Our 1946 Clinic proved self-supporting on this basic fee.

The clinic, operating under the auspices of the University Counseling Board, has as its core of staff members experienced faculty counselors who serve each year as freshman counselors on the DePauw staff. To aid these counselors, and incidentally to widen their own horizons, a staff of visiting counselors is also provided. In the 1946 DePauw Clinic the counseling staff totaled twenty-eight, of which ten were visiting representatives from other colleges and progressive high schools. These visiting counselors, coming with a wide variety of counseling experience obtained in institutions located in three different states, have a distinct leavening influence on the entire counseling process.

The tests used in the 1946 Clinic were as follows: *The American Council on Education Psychological Examination*; the complete *Cooperative English Test* (Single booklet edition); the three *Cooperative General Achievement Tests*, in the fields of Social Studies, Natural Sciences, and Mathematics; the *Robinson-Hall (Ohio State) Tests of Reading Ability*; the *Kuder Preference Record*; the *Strong Vocational Interest Blank*; and the *DePauw Adjustment Inventory*. In addition to this information the counselor had available a complete medical report from the University Health Service and an eight-page autobiographical information blank prepared by the counselee.

The students arrived on Sunday evening and Monday and Tuesday were spent in taking the battery of clinic tests. On Wednesday the students were given the health examination and the counselors met for group sessions and later for smaller committee meetings. At these latter meetings committees of four or five counselors each discussed the cases individually assigned to them, arriving at tentative recommendations on the basis of pooled judgment. The final student-parent-counselor conferences were held on Thursday. Each counselor was given a case load of two counselees. For this reason, in the 1946 Clinic the counselee enrollment was limited to fifty; by increas-

ing the counseling staff for 1947 we have set our new enrollment limit at one hundred.

### *Advantages of the Clinic to the Students*

The values of a comprehensive clinic testing program to the student are perhaps obvious, but we shall list them here for summary purposes. The profile derived from the test battery described above reveals a wealth of basic information helpful to the student. Thus, the profile helps to supply:

1. An analysis of past achievement
2. An analysis of ability level
3. An analysis of vocational interest pattern
4. An analysis of personality traits

From these analyses and their interpretation the student should derive:

5. Help in planning his educational program
6. A considerable amount of vocational guidance
7. Beneficial insight into special deficiencies or special aptitudes that might affect his academic or vocational goals
8. An acquaintance with general college procedures
9. Suggestions as to the type of school best fitted to offer courses suited to his needs
10. A friendly acquaintance with individual faculty members

Many of these potential advantages for the counselees are objectively evaluated in the opinion-survey reproduced below. We can not reproduce the pleasure and enthusiasm exhibited by these counselees; we can only offer testimony that these subjective factors were present to a degree highly satisfying to those of us in charge of the clinic project.

It is apparent that certain counseling problems *cannot* be carried to satisfactory conclusions within the duration of the clinic. Conditions requiring therapy or further education, e.g., personality maladjustments or reading disabilities, can only be identified. Treatment is necessarily a longer term procedure, hence beyond the avowed scope of the clinic. When such issues are involved the clinic staff makes *referrals* to appropriate facili-

ties accessible to the student. Vocational and educational diagnosis and planning are of necessity the primary functions of a program of limited duration.

### *Advantages of the Clinic to the University*

From the standpoint of the university sponsoring this program certain values may be legitimately anticipated. Some of these are more evident than others and the list given below probably does not exhaust all the potentialities. They may be summarized as follows:

1. In making this *service* available to the public a college is building up the best kind of public relations.

2. The sponsoring college profits because its own staff counselors gain much from the opportunity to make these intensive case studies.

3. The local counselors definitely benefit from the association with visiting counselors imported from responsible positions in personnel offices elsewhere.

4. The local counselors become acquainted with some of their counselees for the ensuing year in a manner far superior to that usually obtaining in the typical counselor-counselee relationship.

5. A program of this nature tends to attract particularly earnest students, i.e., those seeking to make the best possible use of their abilities and opportunities. It is reasonable to assume that these students will in turn be attracted to an institution willing to sponsor this counseling service. (The clinic is not designed as a recruiting service, as would be perfectly obvious to anyone in these days of overcrowded colleges.)

6. In certain cases of debatable admission qualifications, a mutually satisfactory answer may be provided by the clinic as to the admissibility of certain candidates.

7. The testing program of the clinic, insofar as it includes the usual entrance examinations, will to some extent relieve the testing load during orientation week.

No effort has yet been directed toward objective evaluation of the points just suggested. Again we must rely on the personal reactions of participants, in this case the counselors and

the university administration. Allegheny has held the clinic as an *annual* feature since 1938. DePauw's first effort was in 1945 and, after a second satisfactory trial in 1946, we have also decided to place the clinic on an annual basis as an integral part of the university program.

Both local and visiting counselors have been vigorous in declaring their pleasure at participation in the clinic. Most of these counselors agree on the values derived, both by the counselees and by the counseling staff. The *unanimity* of approval for this project is so marked as to constitute a relatively rare phenomenon in academic circles.

### *An Opinion Survey of the Participants*

An effort was made to secure objective data evaluating the 1946 clinic from the viewpoint of the students and their parents. We felt it desirable to check the possibility that perhaps the clinic staff was overestimating the apparent enthusiasm with which the program was being received. Accordingly, the questionnaire reproduced below was mailed to the clinic students a few days after their return home. Replies were eventually received from every student. The resultant data indicated in the questions duplicated here are expressed in terms of the *per cent* voting each alternative. Where certain response alternatives were omitted, this has been indicated; this occurred only in some of the parents' section answers.

### OPINION SURVEY ON EDUCATIONAL GUIDANCE CLINIC

An intelligent public relations policy in business procedure involves "customer surveys." Procedures need evaluation in the light of their impression upon the consumer. You were among our "consumers" in the recent Educational Guidance Clinic and we would appreciate your frank and sincere cooperation in completing this questionnaire for our information. Please feel free to vote *any* of the choices indicated; we do not resent honest criticism. Note one section is for *you* and the latter part is for your *parents*. Thanks for your *early* attention to this request and return the questionnaire in the enclosed envelope by July 10th, if possible. Your signature is *optional*. Best wishes for a good summer!

*Student's Section*1. *General Value of the Clinic:*

- 0% (a) Disappointed me extremely; was waste of time  
 5% (b) Not as useful as I hoped; did have some value  
 13% (c) Relatively successful; definitely useful  
 38% (d) Quite satisfactory; value reached my expectation  
 44% (e) Very satisfactory; better than anticipated

2. *Functions of the Clinic:*

(Numbers are per cent)

	Worth- less	Little Value	Moder- ate Value	Valu- able	Very Use- ful
(a) Orientation to col- lege procedures	0	3	8	39	50
(b) Planning educa- tional program	0	3	11	34	53
(c) Vocational Guid- ance	0	5	13	42	39
(d) Analysis of abilities	0	0	11	45	42
(e) Analysis of past achievement	0	3	26	53	18
(f) Acquaintance with faculty	0	3	8	37	53
(g) Acquaintance with new freshmen	0	0	3	26	71

3. *Features of the Clinic:*

(Numbers are per cent)

	Needs Defi- nite Im- prove- ment	About Right	Better than Ex- pected
(a) Ability tests	5	76	18
(b) Achievement tests	3	61	24
(c) Interest tests	5	61	34
(d) Personality test	3	63	34
(e) Health examination	21	63	16
(f) Counseling interviews	3	29	68
(g) Impartiality of advice	0	53	47
(h) Friendliness of staff	0	16	84
(i) Social-recreational functions	3	63	34
(j) Rooming facilities	0	26	74
(k) Meals	8	42	50

*Parents' Section*

(No. %)

Yes ? No Omitted

97	3	0	0	(1a)	Did you expect the Clinic to provide curricular guidance?
91	3	3	3	(1b)	If so, were you satisfied?
94	0	6	0	(2a)	Did you expect the Clinic to provide vocational guidance?
86	0	8	6	(2b)	If so, were you satisfied?
69	6	25	0	(3a)	Did you expect the Clinic to provide orientation to college life?
83	0	0	17	(3b)	If so, were you satisfied?
94	0	6	0	(4)	Do you feel you were sufficiently well acquainted with the purpose of the Clinic?
44	3	50	3	(5)	Would you have liked a general meeting of parents (before the final interview) to explain the testing and counseling procedures more completely?
80	11	6	3	(6)	Do you feel there was enough personal contact with the counselors?
94	6	0	0	(7)	Did the Clinic (as a whole) measure up to your expectations?
100	0	0	0	(8)	Would you recommend the Clinic to other students and parents?

One may well question the final competence of the respondents to pass judgment on certain items above, e.g., Questions 2c, 3a, 3b, etc. We only suggest that how they *feel* about these issues in this clinic situation is probably more significant to them than the possession of technical training to make valid judgments. From the standpoint of underlying emotional satisfactions it is very apparent that the counselees and their parents were highly pleased with their clinic experience. As it now stands we are sure the clinic needs many improvements, but it does seem definitely worthwhile to have been able to present a service generating so much desirable "consumer" good will.

*Summary*

This paper has described a contemporary example of the Educational Guidance Clinic program pioneered by Allegheny



College in 1938. The theoretical values of this procedure as anticipated for the participating counselees and the sponsoring institution have been presented. As objective evidence of clientele reaction we have reproduced an opinion survey carried out among the counselees registered in the 1946 Educational Guidance Clinic at DePauw University.

This type of educational venture has several very commendable qualities, but thus far it has been attempted by only a few scattered schools. Further exploration into this area of concentrated guidance procedures, oriented specifically toward college and educational problems and designed to fit individual cases, will repay considerable investigation on the part of a much larger number of educational institutions. It is a project offering worthwhile returns to all parties concerned, bearing all the signs of a sound educational investment.

#### REFERENCE

- 1 Anderson, Hurst R. "An Educational Guidance Clinic," *Journal of the American Association of Collegiate Registrars*, XVIII (1942), 32-35.



# THE NUMBER RELATIONS SECTION OF THE CARNEGIE MENTAL ABILITY TESTS TREATED AS A POWER TEST<sup>1</sup>

WILEY S. GARRETT

Niles Public Schools, Niles, Ohio

THE *Carnegie Mental Ability Tests*, Number Relations Section (1), has been found by the author to be a very useful instrument in a number of guidance situations. The tests measuring "number relations" include items of both mental manipulation and achievement which make the tests very useful in test batteries where a quantitative aspect of intelligence needs to be measured. This problem-solving aspect of intelligence, involving mathematical concepts, is an important one in guidance because many college courses draw heavily on this quantitative aspect of intelligence for success. Many occupational choices, such as engineering, accounting, and mathematics teaching, are not feasible for those who lack ability in this area. It is an important consideration for all types of technical work and is probably an essential qualification for many workers such as bookkeepers, bank clerks, paymasters, machinists, etc.

The tests when used extensively have two annoyances. First, each of the three subtests has a time limit which the examiner must carefully observe for each test. Secondly, the sharp time limits force the testee to work under excessive pressure. This places a premium on speed and often leaves unsolved many of the items in each part which the testee might have been able to solve had the time limits been more liberal. This has a tendency to leave the testee with a feeling that he

---

<sup>1</sup>The writer wishes to express his gratitude to Dr. William R. Grove, Director of the Division of Psychological Services at the University of Pittsburgh, for his helpful criticisms and valuable suggestions.

has been denied an opportunity to demonstrate his ability completely. This is especially true in the case of the slow but accurate individual.

A solution for the elimination of these two annoyances would be to change the test from a time-limit to a work-limit or power test providing this could be accomplished without injuring the effectiveness of the test.

A review of the literature shows some studies concerning the question of speed versus power. May and Terman (4) obtained an  $r$  of 0.956 for 510 soldiers on the *Army Alpha Test*, when they administered the test with regular time limits and with the time limits doubled. Ruch and Koerth (5), using the *Army Alpha Test* with 122 college freshmen, administered the test using regular, doubled, and unlimited time limits. The obtained  $r$  for regular time versus double time was 0.966 while the  $r$  for regular time versus unlimited time was 0.945. Ruch (6) in another study used the *Terman Group Test of Mental Ability* and the *Stanford Achievement Test*. Using the scores from these tests obtained by giving them with regular time limits and unlimited time limits, he found the following correlations:

Correlations of Regular versus Unlimited Time Limits

<i>Test</i>	<i>N</i>	<i>r</i>
Terman Group Test .....	150	0.960
Stanford Reading Test .....	64	0.968
Stanford Arithmetic Test .....	64	0.976

These studies seem to indicate that regardless of the time limits, whether doubled or unlimited, the relationship remains practically constant. But in these studies the correlations are spuriously high because they depend on a part-whole relationship. The scores obtained in unlimited time or double time in the above studies all have the common element "regular time" plus whatever the subject accomplishes in the additional time. The correlation between scores obtained during regular time and unlimited time is a function of the length of the regular time because as the regular time is lengthened the regular time scores become more similar to scores obtained in unlimited time. In the above studies it might have been more pertinent

to have found correlations between regular time and the increment earned during the added time. This would have given a truer relationship between regular time scores and unlimited time scores.

### *The Experiment*

The specific problem of this study will be to investigate the difference in scores on the *Carnegie Mental Ability Tests*, Number Relations Section, obtained by administering the tests in such a manner that they can be scored as both power and time-limit tests.

The subjects for the experiment were 155 students of McKinley High School located at Niles, Ohio. The sample consisted of 85 girls and 70 boys all enrolled in the 12th grade. Three tests were administered to this group on November 7, 1945: The *Ohio State Psychological Examination*, Form 21 (7), was given in the morning, and the *Carnegie Mental Ability Tests*, Number Relations Section (1), and the *New Stanford Arithmetic Test*, Primary Form Z (3), were given in the afternoon. The primary form of the *Stanford Arithmetic Test* was chosen to insure a measure of those students who might have a very limited ability in arithmetic.

The Ohio State Test and the *Stanford Arithmetic Test* were administered in the manner prescribed in the *Manual of Directions* for these tests. The *Carnegie Mental Ability Tests*, Number Relations Section, was administered using the following directions: After the subject had worked on the first subtest for the time limit prescribed in the *Manual of Directions*, he was instructed to draw a line on the answer sheet under the last problem worked. This was repeated for the second and third subtest. The student was then given a red lead pencil and asked to turn back to the first subtest and instructed, "You may have additional time to finish this test. Do all your writing with your red pencil. If you want to change an answer above the line you have drawn, do not erase but enter your corrected answer in red beside your answer written in pencil." Time was called when the examiner estimated that about 95% of the group had reached their limit. This procedure was re-

peated for the second and third subtests. Table 1 shows the prescribed time, the time added, and the total testing time for each of the three subtests:

TABLE 1  
*Time Allowed for Tests 4, 5, and 6 of The Carnegie Mental Ability Tests*

Name of Test	Prescribed Time Limit (Minutes)	Time Added Each Test (Minutes)	Total Time (Minutes)
Test 4, Number Series (30 items) ..	9	9	18
Test 5, Calculation (40 items) .....	11	9	20
Test 6, Mathematical Reasoning (35 items) .....	15	12	27
Total (105 items) .....	35	30	65

This technique yielded three scores:

1. A prescribed time-limit score, which was the number of problems solved written with the ordinary pencil down to the line drawn on the answer sheet. This score is in every way comparable to scores usually obtained on these tests when they are given in the manner prescribed in the *Manual of Directions*.

2. A power score, which was the number of problems solved for the entire test written both in pencil and in red. If a problem had two answers, red and black, the red answer was considered in the power scoring only.

3. An increment score, which was the number of problems solved written with the red pencil.

To avoid confusion, hereafter, the *Carnegie Mental Ability Tests*, Number Relations Section's scores will be referred to as:

CNR-Time-Limit (scores obtained during prescribed time limits)

CNR-Power (scores obtained during prescribed time limits plus additional time)

CNR-Increment (scores obtained during additional time only)

### *Results*

The medians, semi-interquartile range, and skewness of the distributions are given in Table 2. In each case, except for the *Stanford Arithmetic Test* where scores are reported as age and grade scores, the data were computed from the norms given in

the *Manual of Directions* for that test and compared with the results obtained in this study. The Table indicates that the distributions of each of the tests, with the exception of the *Stanford Arithmetic Test*, show a tendency for the scores to cluster toward the lower end of the scale. The negative skewness of the *Stanford Arithmetic Test* might be expected because the primary form of that test was used. The  $Sk/\sigma sk$  for all the distributions, except for CNR-Time-Limit, are above three and, therefore, significant. This indicates that these distributions are positively skewed to the right. The  $Sk/\sigma sk$  of .92, obtained for CNR-Time-Limit, denotes that there are 83 chances in 100 that the obtained  $Sk$  of 1.08 is significant. The departure of this frequency distribution from normality, therefore, while fairly large is not marked.

TABLE 2  
*Medians, Semi-Interquartile Range, and Skewness for the  
Tests Used in the Study*

Name of Test	<i>N</i>	Median	<i>Q</i>	<i>Sk</i>	$\sigma sk$	$\frac{Sk}{\sigma sk}$
CNR-Time-Limit . . .	155	26.66	8 0	1.08	1.17	.92
Manual of Directions . . .	500	29.00	7 5	3.50	.68	4.8
Ohio State Test . . . .	155	49.54	15.5	8.47	1.56	5.4
Manual of Directions . . .	13,380	50.00	17.5	12.75	.92	13.8
Stanford Arithmetic . . . . .	155	102.79	102.1	-6.60	1.44	-4.5

The difference between the median for CNR-Time-Limit and the median quoted in the *Manual of Directions* for that test is 2.34 and the  $D/PE_D$  is 2.6. The difference between the median for the Ohio State Test and the Median quoted in the *Manual of Directions* is .46 and the  $D/PE_D$  is .36. This indicates that these differences are not too great for all reasonable purposes in this experiment and that the results for these tests for Niles McKinley Seniors do not differ widely from the scores usually obtained and that this experimental group forms an adequate sample of the 12th grade.

Table 3 gives the product-moment coefficients of correlations found by intercorrelating the variables. It should be noted that the test scores for all combinations of the CNR

Tests and the *Ohio State Psychological Examination* are total raw scores, no account being taken of the fact that these tests yield scores on two or more subtests. The *Stanford Arithmetic Test* score is an equated average.

The correlations shown in Table 3 are all positive, they are statistically reliable, and they range from fairly high to high. This would indicate that the same general function is being measured but that specific test content is different, except for the combinations CNR-Time-limit versus CNR-Power (.96) and CNR-Increment versus CNR-Power (.90) which are quite high indicating very little change in rank in these distributions. These findings are similar to those of May and

TABLE 3  
*Coefficients of Correlations Between Variables, Means and Standard Deviations of Each Variable*

Variable	CNR- Time- Limit	CNR- Power	Ohio State	Stan. Arth.	Mean	S.D.
CNR-Increment .....	.77	.90	.46	.72	15.43	6.36
CNR-Time-Limit .....	..	.95	.55	.75	27.22	10.84
CNR-Power .....	..	..	.56	.79	42.77	16.15
Ohio State .....	..	..	..	.56	54.61	22.12
Stanford Arithmetic .....	..	..	..	..	99.29	13.40

Terman (4), Ruch and Koerth (5), and Ruch (6) who found  $r$ 's ranging from .94 to .96 between regular time versus doubled or unlimited time. But in all of these studies,  $r$ 's are spuriously high because of the common element regular time plus whatever the subject accomplishes during the added time. In the case of CNR-Time-Limit versus CNR-Power the common element is regular time limit which is part of the Power score and for CNR-Increment versus CNR-Power it is the increment which is also part of the Power score.

This spurious element was factored out by correlating CNR-Time-Limit with CNR-Increment. The obtained  $r$  of .77 is very high and indicates a fairly close relationship between the *Carnegie Mental Ability Tests*, Number Relations Section, given under regular time limits and the increment the student earned during the added time. The prediction index from



this  $r$  denotes that a score from one of these variables might predict the score on the other about 36% better than a guess. An analysis of the scattergram for these two variables shows that 73 of the 155 cases were located within the range of  $\frac{1}{2}$  sigma, that is, a pupil's score on CNR-Time-Limit, within this range, received a similar rank on CNR-Increment; 72 of the 155 cases were located approximately 1 sigma away from the predicted position; and only 10 cases were misplaced by approximately 2 sigmas. This is evidence of a relationship which is far superior to chance.

The correlations between the Carnegie Test combinations and the *Stanford Arithmetic Test* (number relations) with the Ohio State Test (verbal factor) range from .46 to .56. This shows some relationship between these variables and indicates that scores on the Ohio State Test can predict scores on the Carnegie Tests and the *Stanford Arithmetic Test* with equal facility, although the prediction index from these  $r$ 's is only 16% and 17%, respectively, better than a guess.

The  $r$ 's obtained between the Carnegie Test combinations and the *Stanford Arithmetic Test* (ranging between .72 and .79) are high, indicating that these two types of tests measure somewhat the same function.

A comparison of the means and standard deviations (Items 3 and 4, Table 4) will give some indication of the effect of increasing the time limits for CNR-Power. Approximately, the upper 1/6 of the seniors were located above a score of 38.06 on CNR-Time-Limit while the lower 1/6, approximately, were located below a score of 26.62 on CNR-Power. The poorest 1/6, after the time was increased, scored within 11 score points of the best 1/6 working under prescribed time limits.

The *Manual of Directions* for the *Carnegie Mental Ability Tests* (2) reports a reliability coefficient of  $.888 \pm .008$  for the sum of tests, 4, 5, and 6 (CNR-Time-limit). This reliability was determined by correlating odd with even items of the tests. The  $r$  calculated in the same manner for CNR-Power for McKinley High School Seniors was  $.951 \pm .005$ . The difference between these  $r$ 's (.888 and .951) is .063 and is a significant difference as shown by the *D/P.E.* Difference of 6.7 in favor

of the tests given under increased time limits. But the computed reliabilities for CNR-Time-limit and CNR-Power (Item 6, Table 4) for the Niles McKinley Seniors are .931 and .951, respectively. The difference between these  $r$ 's is .020 and the *D.P.E.* Difference is 2.5 in favor of the tests given under increased time limits. Though the difference is not significant, it is large enough to indicate that increasing the time limits of the test tends to increase their reliability as far as the Niles McKinley sample is concerned.

The standard error of measurement (Item 7, Table 4) for CNR-Time-limit is 2.85 and for CNR-Power it is 3.57. This

TABLE 4  
Statistics for CNR Time-Limit and CNR-Power

Item	CNR-Time-Limit	CNR-Power
1. Number .....	155	155
2. Time .....	35 min.	65 min.
3. Mean .....	27.22	42.77
4. Standard Deviation .....	10.84	16.15
5. Std. Error of Mean .....	0.87	1.28
6. Coefficient of Reliability ..	0.931 $\pm$ .007	0.951 $\pm$ .005
7. Std. Error of Measurement ..	2.85	3.57
8. Std. Error of Estimate .....	3.17	4.73
9. Percentiles		
90 .....	48	66
75 .....	38	54
50 .....	29	43
25 .....	23	29
10 .....	17	23

is only .26 and .22 of a sigma respectively and presents fairly good evidence that increasing the time limits of the tests makes very little difference in the two distributions.

The standard error of estimating scores under CNR-Time-limit from those earned under CNR-Power is 3.17. Similarly, the standard error of estimating scores under CNR-Power from those earned under CNR-Time-limit is 4.73 score points. These measures are only .29 and .33 of a sigma of their respective distributions and again indicate that increasing the time limit distorts the measurement between the two distributions very little.

Percentiles computed for CNR-Power (Item 9, Table 4), when compared with the percentiles given in the *Manual of*

*Directions* for CNR-Time-limit, show that pupils taking the tests under increased time limits earn scores of 14 points above CNR-Time-limit at the median, 18 points higher at the 90th percentile, and 6 points higher at the 10th percentile. But the very high correlation between these two distributions (.956) indicates very little change in rank between the two tests. Evidently, increasing the time limits does not change the relative position of the student in relation to the group in any significant fashion.

### *Summary and Conclusions*

In this study the difference in scores for the *Carnegie Mental Ability Tests*, Number Relations Section, was investigated by giving the tests to 155 McKinley high-school seniors under prescribed time limits and under increased time limits. The results of this study show no significant difference in the measuring quality of the tests when the time limits are increased. This is clearly shown by the extremely high correlation of .956 obtained when CNR-Time-limit was correlated with CNR-Power. This relationship is still high, an obtained  $r$  of .772, when the scores of CNR-Increment are correlated with CNR-Time-limit. Furthermore, the standard error of measurement for an obtained raw score is 2.85 for CNR-Time-limit and 3.57 for CNR-Power. This is only .26 and .22 of a sigma respectively and indicates that increasing the time limits distorts very little the measurement between the two distributions. The standard error of estimating scores under CNR-Time-limit from those earned under CNR-Power is 3.17 and for estimating those under CNR-Power from those earned under CNR-Time-limit it is 4.73. These measures are only .29 and .33 of a sigma for the respective distributions and indicate that increasing the time limits make very little differences in the two distributions.

If increasing the time limits of the *Carnegie Mental Ability Tests*, Number Relations Section, has very little effect on the rank position of the student, then changing the test from one designed to measure speed and accuracy to a test of power will not materially affect the original measuring quality of the test. Moreover, the test will have the following advantages:

1. It will relieve the examiner from timing each subtest, freeing him to work on other tasks.
2. It will moderately increase the reliability of the tests.
3. It will relieve the testee from working under excessive pressure and will give him the satisfaction of knowing that he has reached the level of his accomplishment.
4. It will, if standardized as a power test, provide a measuring instrument for "Number Relations" which can be used as either a time-limit or a power test depending on the needs of the administrative situation.

From the results of this study the author believes that the modification of the time limits of the *Carnegie Mental Ability Tests*, Number Relations Section, is a valid substitute for the original time limits of the tests and that this revised test could be very useful as a guidance instrument if standardized as a power test.

#### REFERENCES

1. Cleeton, Glen U. *Carnegie Mental Ability Tests*. New York: Houghton Mifflin Company, 1931.
2. Cleeton, Glen U. *Manual of Directions, Carnegie Mental Ability Tests*. New York: Houghton Mifflin Company, 1931.
3. *New Stanford Arithmetic Test*, Form Z. Yonkers-on-the-Hudson: World Book Company.
4. "Psychological Examining in the United States Army." *Memoirs of the National Academy of Science*, XV (1921), 416.
5. Ruch, G. M. and Koerth, W. "Power versus Speed in Army Alpha." *Journal of Educational Psychology*, XIV (1923), 193-208.
6. Ruch, G. M. "The Speed Factor in Mental Measurement." *Journal of Educational Research*, IX (1924), 39-45.
7. Toops, Herbert A. *Ohio State Psychological Examination*, Form 21. Columbus, Ohio: Ohio State University.

# PREDICTION OF THE DIFFICULTY INDEX OF OBJECTIVE-TYPE SPELLING ITEMS

C. W. HARRIS

University of Chicago

ONE of the problems that arises in the construction of various kinds of achievement examinations is that of adjusting the individual items—and consequently the entire examination—to the desired difficulty level or levels. Ordinarily, the examiner attempts to build items whose distribution of difficulty indexes will approximate the desired distribution of difficulty; to achieve this, he often must rely largely on his own previous experience in building items of this type and on his guesses or “hunches” regarding the probable difficulty of particular items. Presumably the examiner does anticipate the distribution of difficulty indexes that will be appropriate for the particular examination he is building<sup>1</sup> and attempts to match that distribution with the items that he is preparing.

It is of interest to raise the question of whether or not the difficulty of items, with respect to the particular population for whom they are intended, can be predicted. This is to ask, in effect, two questions: (1) Is the difficulty index “reliable,” i.e., consistent from sample to sample? and (2) If this index is consistent from sample to sample, does it have a positive and statistically significant relationship with any variable that is ordinarily available to the examiner at the time of item construction? These questions have been answered for what is probably one of the simplest types of item to construct; namely, items testing the recognition of misspelled words. These items are in the form:

---

<sup>1</sup> For a recent discussion of one aspect of the problem of what constitutes an “appropriate” distribution of difficulty, see Gulliksen, Harold, “The Relation of Item Difficulty and Inter-Item Correlation to Test Variance and Reliability,” *Psychometrika*, X (1945), 79-91

1. hallucination
2. obstreperous
3. changing
4. unmitigated
5. none misspelled

The subject's task is to determine which one, if any, of the four words given is misspelled and to indicate his answer by recording the number of that word; if he believes that none of the four words is misspelled, he records number 5. In the example above, number 3 is obviously the correct answer. For the various sets of items of this type used in this study, the practice was followed of using each of the five responses, in approximately equal proportions, as the correct response.

The question of whether or not the difficulty index of items such as these remains constant from sample to sample was answered by determining the product-moment coefficient of correlation between the difficulty indexes derived from two different samples of subjects. Two different sets of items of this type were used in order to provide more than one datum as a basis for the answer. Set I includes 40 items; these appear in Forms A and B (twenty items in each) of Test I of the Tests of General Educational Development (High School Level). Set II includes twenty-six items; these were prepared for tryout for inclusion in new forms of this test. Each of the two sets of items was administered to two samples of subjects. Set I was administered to 270 high-school seniors and also to 200 veterans who were taking the entire series of the General Educational Development Tests. Set II was administered to 124 high-school seniors and also to 106 veterans. The difficulty index for each item was computed by determining the per cent of the sample that chose the correct answer. Two difficulty indexes were computed for each item—one derived from the sample of high-school seniors and the other from the sample of veterans. Correlation coefficients between difficulty indexes for each of the two sets of items are presented in Table 1. (It will be noted that the number of cases on which each of these coefficients is based is the number of *items* rather than the number

TABLE 1  
*Relationship Between Difficulty Indexes Derived from Two Different  
 Samples of Subjects*

SET I				$r = .905$
(n = 40 items)				
	<i>N</i>	<i>Mean</i>	<i>SD</i>	
High school seniors . . . . .	270	61.72%	14.29%	
Veterans .. . . .	200	54.68%	15.13%	
(For $n = 40$ , an $r$ greater than .403 is significant beyond the 1% level)				
SET II				$r = .827$
(n = 26 items)				
	<i>N</i>	<i>Mean</i>	<i>SD</i>	
High school seniors ...	124	59.31%	15.57%	
Veterans .. .. .	106	48.42%	17.58%	
(For $n = 26$ , an $r$ greater than .496 is significant beyond the 1% level)				

of subjects.) Both of these correlations are reasonably high (.90 and .83), are positive, and are statistically significant. They indicate that the difficulty index derived from one sample is linearly related to the index derived from another sample. The means indicate that the items tend to be easier (i.e., a greater per cent passes) for the high-school seniors, but the correlation coefficients indicate that this tends to be a consistent rather than a random difference, with respect to each item.

It appears, then, that the difficulty index of items such as these has a characteristic that is necessary if it is to be predicted—that of consistency from sample to sample. An attempt was therefore made to determine whether or not the difficulty index could be predicted. The variable that was used in this attempt will be called a scale value. This scale value was derived from data presented by Simmons and Bixler.<sup>2</sup> Simmons and Bixler give scale values (in per cent) for a large number of words commonly used in spelling tests and lessons; each scale value represents the per cent of students able to spell the word correctly. In this study, the twelfth-grade scale value is used throughout; Simmons and Bixler report that these values were derived from

<sup>2</sup> Simmons, Ernest P. and Bixler, Harold H. *The New Standard High School Spelling Scale*. Atlanta: Turner E. Smith and Company, 1940, 64 pp.

"200 or more" students.<sup>3</sup> For the first two sets of spelling items (Set I and Set II) the *average* scale value of the four words included in the item was used as the scale value for the item.<sup>4</sup> These items had been built without the thought of including in a particular item only words of the same scale value. Later, this idea was hit upon, and two additional sets of items were built according to the scheme of including in any item only words of the same scale value. For the first test of the practicability of predicting the difficulty of items of this type, however, the average scale value for what may be called heterogeneous items—i.e., items made up of words of different scale values—was used. The product-moment coefficients of correlation between this average scale value and the difficulty index are given in Table 2. As would be expected, the difficulty index derived from the one sample correlates with the scale value to approximately the same extent as does the difficulty index derived from

TABLE 2  
*Relationship Between Difficulty Index and Average Scale Value of  
Heterogeneous Spelling Items*

SET I			
(n = 40 items)			
	Mean	SD	
Difficulty index (veterans) .....	54.68%	15.13%	$r = .241$
Average scale value .....	85.55%	5.52%	
Difficulty index (high school) .....	61.72%	14.29%	$r = .256$
Average scale value .....	85.55%	5.52%	

---

SET II			
(n = 26 items)			
	Mean	SD	
Difficulty index (veterans) .....	48.42%	17.58%	$r = .067$
Average scale value .....	81.31%	7.34%	
Difficulty index (high school) .....	59.31%	15.57%	$r = -.017$
Average scale value .....	81.31%	7.34%	

<sup>3</sup> *Ibid.*, page 32.

<sup>4</sup> For a very few items it was necessary to average three, rather than four, scale values, since one of the words did not appear in the Simmons-Bixler list.



the other sample. For Set I, the correlation between average scale value and difficulty index is relatively low, but positive. For Set II, the correlation differs little from zero.

The answer to the question of whether or not the average scale value is useful in predicting the difficulty index of heterogeneous spelling items is that very little practical control can be secured under these conditions. Inspection of the standard deviations of the average scale values, given in Table 2, suggests that the low relationships may be primarily a function of the restricted range of the one variable. It is obvious that in averaging scale values of four words in order to secure a scale value for the item, the effect is to produce a variable whose range is restricted in comparison with the range of scale values of all the words used in the items. This would not be true, of course, if each item were made up only of words of the same scale value. In order to provide a second test using these heterogeneous items, the scale value of the word that is misspelled in each item was assigned as the scale value for the item. For those items in which no word is misspelled, the average of the scale values of the four words was used. The correlation between this scale value, which will be called the *response* scale value to differentiate it from the average scale value, and the difficulty index derived from the sample of high-school seniors was computed for Set I and for Set II. These correlations are presented in Table 3. Both coefficients are positive; however, neither is very high. The response scale value has, in each case, a standard deviation roughly twice that of the average scale value; the coefficient for Set I, however, decreases with this increase in the range of the scale value variable, whereas the coefficient for Set II changes sign and increases. Again, it would appear that very little practical control of the difficulty index can be secured by using a scale value of this sort as a predictive variable.

These investigations of the relationship between scale value and difficulty index for heterogeneous items were followed by the development of what will be called homogeneous items and the study of this relationship for them. As additional spelling items were being built for tryout, preparatory to constructing new forms of the General Educational Development Tests, an

attempt was made to make additional use of the Simmons-Bixler scale. The idea of including in any item only words of the same scale value suggested itself as a promising technique. For items of this type, the average scale value and the response scale value are identical. In order to provide a test of the relationship between scale value and difficulty index under these conditions, two new sets of items were constructed and tried out. Set III consists of thirty-five homogeneous items, and was administered to 113 high-school seniors. Set IV consists of forty homogeneous items, and was administered to 105 students. By design, a comparatively narrow range of scale values

TABLE 3  
*Relationship Between Difficulty Index and Response Scale Value of  
Heterogeneous Spelling Items*

Set I			
(n = 40 items)			
	Mean	SD	
Difficulty index (high school) . . . . .	61.72	14.29%	$r = .126$
Response scale value . . . . .	85.03%	9.72%	
Set II			
(n = 26 items)			
	Mean	SD	
Difficulty index (high school) . . . . .	59.31%	15.57%	$r = .299$
Response scale value . . . . .	76.68%	13.61%	

was used in building Set III, and a wide range was used in building Set IV. The difference in range of scale values for the two sets is apparent from Table 4, in which the coefficients of correlation between scale value and difficulty index for these two sets of items are given. The correlation for Set III, which has a very narrow range of scale values, is about as high as any of the correlations found for heterogeneous items, despite the fact that this range of scale values for Set III is less than that for either Set I or Set II. The correlation for Set IV, which has a fairly wide range of scale values, is reasonably high. It would appear, then, that a fairly satisfactory correlation between difficulty index and scale value can be secured if homogeneous items

TABLE 4  
*Relationship Between Difficulty Index and Scale Value for  
 Homogeneous Spelling Items*

SET III		
(n = 35 items)		
	Mean	SD
Difficulty index (high school) . . . . .	71.97%	17.95%
Scale value . . . . .	92.71%	3.66%
(For n = 35, an r greater than .430 is significant beyond the 1% level)		
SET IV		
(n = 40 items)		
	Mean	SD
Difficulty index (high school) . . . . .	60.08%	20.24%
Scale value . . . . .	74.35%	12.10%
(For n = 40, an r greater than .403 is significant beyond the 1% level)		

are used and if the distribution of scale values for these items has a fairly wide range.

The technique developed in this study may have value for other examiners who are engaged in building items of this type. Within practical limits, prediction of the difficulty index from the scale value for homogeneous spelling items is possible. The value of  $b_{xy}$  for predicting difficulty index ( $x$ ) from scale value ( $y$ ) is 1.30 when the data for Set III are used, and is 1.23 when the data for Set IV are used. It would seem reasonable to proceed, at least tentatively, on the assumption that the deviation of the difficulty index from its mean will approximate 1.25 times the deviation of the scale value from its mean.

The extension of the use of homogeneous items is suggested. Vocabulary items, items testing the ability to perform numerical operations, and the like, should be amenable to this technique of construction. If they are, it would be of interest to determine whether or not one could secure similar control of their difficulty indexes.



## STUDIES OF POPULARITY IN COLLEGE: II. DO DORMITORY ARRANGEMENTS AFFECT POPULARITY?

JEAN WAID REILLY AND FRANCIS P. ROBINSON

Ohio State University

COUNSELORS charged with dormitory arrangements may often wonder if locating a student's room on a busy corridor or even having the student room with a popular person may not aid his or her popularity. Or conversely, one may wonder if assigning a student to some far room may tend to isolate him or her socially. In light of the importance with which students view social status and of the injurious effects on personality development of frustrating social isolation, it seems obvious that some data on these questions should be obtained.

Some preliminary answers are provided in a study made in a girls' dormitory at Ohio State University. The one hundred and sixty-three girls in the dormitory were asked to list for eight campus situations the girls on the campus whom they would most like as companions; this simple sociometric test thus provided by secret ballot a measure of each girl's popularity.<sup>1</sup> With such scores for each girl it is possible to study their relationship to various dormitory conditions.

The dormitory has the following characteristics which form elements for this study.

1. There are four floors, all are of similar plan except the first. Although an elevator is available at the back of the dormitory, most travel is by the stairway. Thus, some insight may be obtained into the relative popularity of girls living on the ground floor near the lounges, on the top floor at the end of the stairway, and on the intervening floors by which the girls on the floors above must pass.
2. The second, third, and fourth floors each have two "large" (16 or 17 girls each) and two "small" (6 or 8 girls each)

---

<sup>1</sup> The first article in this series discusses the details of the construction and administration of the test. Jean W. Reilly and Francis P. Robinson, "Studies of Popularity in College: I. Can Popularity of Freshmen be Predicted?" *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VII (1947), 67-72.

corridors, one of the long corridors being quite far distant from the stairway. These six large and six small corridors permit a judgment on any tendency for one size to be preferable to the other.

3. Not only is it of interest to study the average popularity of the girls in a given corridor but also to find out how much they tend to select each other as friends rather than to look elsewhere. This tendency for a group to form a friendly unit probably provides its members with greater feelings of belonging and of being socially secure than would an equal number of widespread friends; this characteristic of a corridor is here called the "cohesiveness" of the group.
4. A fourth element of dormitory arrangement is the effect of a student's popularity on her roommate.
5. Finally, some data can be obtained as to the effect of differences in race, religion, sorority membership and upperclass status on the popularity of girls.

In general, floors II, III, and IV had quite similar types of girls and, as indicated earlier, had similar floor plans. The average number of girls on each of these three floors was 45; the distribution of minority groups was almost identical, and was as follows: sorority members, 38%; upperclassmen, 42%; Jewish girls, 21%. On floor I, the floor plan was smaller, the number of girls fewer (29), and the distribution of minority groups somewhat different, i.e., sorority girls, 45%; upperclassmen, 45%; Jewish girls, 10%; and Negro girls, 7%.

1. The distance up the stairs had little, if any, effect on popularity. The average and the standard deviation of the popularity scores for each floor are given below:

	<i>Average</i>	<i>S.D.</i>
I .....	36.0 .....	18.6
II .....	41.7 .....	22.9
III .....	39.7 .....	23.2
IV .....	40.8 .....	20.2

Thus, girls on the fourth floor were as popular as those on other floors. The only one of these differences which is statistically significant ( $CR = 3.2$ ) is between the averages for floors I and II.

The presence of the two Negro girls on the first floor may be one factor in this floor difference but it does not seem to be the only reason. For instance, if these two students are omitted,

the average popularity for this floor (37.3) is still below that of the other floors. The lower popularity of the girls on the first floor may also be due to the fact that most students think of this floor as a lounge and as "on the way" to the other floors. However, it must be noted that some of the most popular girls lived on the first floor, so it cannot be assumed that living on the first floor will necessarily cause unpopularity.

It seems evident, then, that in typical college dormitories, the floor level on which a girl has her room will have little effect on her popularity with the possible exception that girls in rooms on the ground floor off from lounging facilities may be slightly handicapped.

2. The position of a girl's room, whether it be on a large or small corridor, was found not to be a factor determining her popularity. The average popularity of each type of corridor was about the same (39.3 vs. 40.9) and the per cent of isolated students on these different types of corridors was the same. In addition, it is interesting to note that the short and the long corridors have both highly accepted girls and, at the same time, unaccepted ones. This similarity of pattern suggests that corridor size does not foster popularity nor unpopularity. Consequently, placing a girl on either size corridor will not necessarily help her to become popular.

3. On the other hand, is there any relationship between corridor size and the tendency for its occupants to choose each other as companions? For one personnel goal is the promotion of congenial groupings among students. The extent to which students on a given corridor selected each other is here termed the "cohesiveness score" for that corridor. The average cohesiveness score for all the large corridors was 40% while the average for all of the small ones was only 28%. Furthermore, the cohesiveness of only one of the small corridors equaled the average for the larger ones. It will be recalled, however, that girls on the small corridors were as popular even though their friends were not as frequently in the same corridor.

4. Some evidence was obtained that merely living close to a popular person does not guarantee popularity for the associate. That is, in three instances some of the least popular

girls were roommates to some of the most popular girls even though they had been roommates for five months. These findings would tend to indicate that popularity cannot be controlled simply by roommate assignment.

5. Finally, some observations may be made on the relation of minority groupings to popularity even though space limitations do not permit a complete presentation of data. (1) The scores for the two Negro girls were 17 and 19 as compared to the average for their floor of 36. Even though these scores are below the average, the interesting point is the degree to which the girls did receive nominations. (2) Girls belonging to the same sorority sought one another as friends no matter where they lived within the dormitory. This "bond," however, existed only between members of the same sorority and did not carry over to other affiliated girls. (3) Jewish girls looked to one another as friends regardless of how far apart their rooms happened to be. But this same tendency did not seem to carry over to the Protestant students. (4) Upperclass students tended to choose more girls outside the dormitory as friends than did the underclass girls.

In brief, it has been shown that the location of a girl's room within the dormitory as well as mere proximity to popular persons has little effect on a girl's popularity. The dormitory counselor, therefore, need not concern herself, particularly, with an attempt to assign room locations according to the probable social acceptability of the girls. She needs to use her efforts to place individuals together who will probably be congenial and who will enjoy one another's companionship. On the other hand, these findings permit a counselor to reassure a girl who has a room in a corner on the top floor that the location of her living quarters will not affect her popularity.

If a counselor is to aid a girl in her efforts to be socially acceptable, she must emphasize personal development to the girl. She should indicate to the girl that popularity depends upon her own characteristics, ones over which she has control, rather than upon such impersonal matters as the location of her room within the dormitory.



# A STUDY OF THE EFFECT OF CONFORMITY TO SOCIAL EXPECTANCY ON EVALUATIVE ATTITUDES

ROBERT P FISCHER

University of Florida

and

AVONNE L ANDREWS

University of Illinois

ATTITUDE tests frequently are used to evaluate specific courses of study. As such, they can be employed to assay the effects of instruction upon opinion or attitude. However, they cannot be used uncritically since they demand considerable honesty on the part of students. Because the results of attitude tests may be used to form educational policy it is especially important to know whether they reflect actual opinion or whether they merely reflect conformity to social expectancy.

This paper reports a study of the effect of conformity to social expectancy as shown by the responses of a group of freshman women on the *Allport-Vernon Study of Values test* (1). The Study of Values<sup>1</sup> test was used in connection with an evaluation of a curriculum in general education at the University of Illinois.<sup>2</sup>

## *Procedure*

The Study of Values test was given a total of three times to the same group of 93 freshman women. It was initially given

---

<sup>1</sup> The *Allport-Vernon Study of Values* test measures the relative dominance of six basic interests in personality. These six interests are the Theoretical, Economic, Aesthetic, Social, Political, and Religious. They are based on Spranger's *Types of Men*, a study which advanced the thesis that the personalities of men could be determined by the values which they considered important (5).

<sup>2</sup> The Division of General Studies is a four-year program in general education open to both men and women in the College of Liberal Arts and Sciences. It consists of seven courses, each one year in length, plus a divisional system of major concentration rather than a departmental system. Because of the nature of the prescribed work in the Division of General Studies, students electing the program tend to have relatively homogeneous educational interests (4).

at the end of the first week of the students' first semester in college. This will be referred to as control test one. It was given the second time at the end of their first semester in college. This will be referred to as control test two.

The students were given the test for a third time immediately after control test two, with the following instructions: "You have taken this test by indicating your personal preferences on the items. Now we would like to have you take the test indicating the way you think you *should* respond to the items. For example, you might *personally* feel that because of the aggressive and self-assertive nature of man, the abolition of war *is* an illusory ideal, but at the same time you might think that you *should believe* that the abolition of war *is not* an illusory ideal." This third test will be referred to as the experimental test since it was the test on which the conditions of administration were purposely altered.

The instructions were intentionally vague as to what was meant by the words, "*should believe*." No attempt was made to determine the manner in which the students interpreted these words, but it is probable that these words were given different interpretations by different students. The important thing was to determine whether the students had marked the test the first two times, controls one and two, in a way, which they felt would be expected of them or whether they had marked them as they actually believed they felt about the items. For this purpose it was thought best to avoid any special set and to let the students interpret the words "*should believe*" in their own way.

### Results

Table 1 compares the results on the two control tests. From this table it may be seen that the mean scores on the six values were virtually the same on both of the control tests. The results on the two control tests did not differ significantly<sup>3</sup> despite the fact that the control tests were separated by an interval of three and one-half months, a period of time which comprised the students' first semester of college work.

<sup>3</sup> The term significant as used in this study refers to a difference between means at least three times greater than the standard error of the difference.

TABLE 1

*Means, Differences between Means and Critical Ratios of the Differences on the Allport-Vernon Study of Values Given at the Beginning and the End of the First Semester in College*

Value	Beginning of semester	End of semester	Difference between means	Critical ratios of the difference between means
Theoretical . . . .	25 423	25 284	139	.149
Economic . .	24 553	24 300	.253	.333
Aesthetic . . . . .	29 536	31 154	-1.618	-1 518
Social . . . . .	33 366	32.809	557	.719
Political . . . . .	30 333	30 326	.007	.009
Religious . . . . .	36 544	36 082	462	.368

Table 2 compares the results obtained on control test two with those obtained on the experimental test. From this table it may be seen that the scores on the first three values, namely Theoretical, Economic, and Aesthetic, were significantly different on the two tests. The Theoretical and Economic scores were significantly higher on the experimental test than they were on the second control test while the Aesthetic scores were significantly lower on the experimental test. These results tend to indicate that the students felt they should be less favorable to Aesthetic values than they had been on the control tests and that they should be more favorable to Theoretical and Economic values than they were when they took the test in the regular manner.

The students scored higher on the Religious value when they marked the tests as they thought they "should believe" than when they marked them in the usual way. They scored lower

TABLE 2

*Means, Differences between Means and Critical Ratios of the Differences on the Allport-Vernon Study of Values when Students Respond to the Items as They Believe and as They Think They Should Believe*

Value	As students think they believe	As students think they should believe	Difference between means	Critical ratios of the difference between means
Theoretical . . . .	25.284	33 003	-7 719	-8 151
Economic . . . . .	24 300	26 972	-2 672	-3 121
Aesthetic . . . . .	31 154	20.377	10 777	9 404
Social . . . . .	32.809	30 783	2 026	2 435
Political . . . . .	30 326	29 369	957	1.149
Religious . . . . .	36.082	39.453	-3 371	-2 535

on the Political and Social values on the experimental test than they scored on the control tests. While none of these last three differences were statistically significant as defined in footnote 3 (Supra), the differences on the Social and Religious values were very close to being statistically significant and indeed would have been by less demanding statistical criteria, criteria which are perfectly acceptable to many statisticians.

Despite the fact that the two control tests were separated by an interval of three and one-half months, which comprised the students' first semester in college, the relative uniformity of scores on the two control tests was to be expected in view of the findings of other investigators (2, 3). It seems reasonable to assume that there was something different in the students' motivation when they took the test under the experimental conditions, in view of the fact that the students scored differently on the experimental tests than they did on the control tests. Since the experimental test actually differed from the control test only in the instructions, it seems safe to conclude that the instructions were responsible in large part for the differential responses.

While these findings do not obviate the possibility that the students marked the tests the first two times as they thought they should, they do make such a conclusion difficult to support. Of course, it could be argued that they took the tests the first two times as they thought they should and changed on the third test since conformity to social expectancy was specifically requested. Such an argument loses much of its force when it is remembered that while the control tests were separated by an interval of three and one-half months, the second control and the experimental test were separated by but a matter of moments. Had they responded in any but a frank way the first time it seems doubtful that they would have responded the same way three and one-half months later for such a condition would require considerable memory.

The total amount of change from the second control to the experimental test was correlated with the scores the students made on the *American Council on Education Psychological Examination*. The correlation thus obtained was .179. Such

a low correlation, though positive, does not reveal any significant relationship between the tendency to resist conformity to social expectancy and intelligence. Such a relationship is not surprising for while superior students might be expected to be more honest and to have more opinions of their own, it is also possible that they are more amenable to social demands.

It seems reasonable to conclude on the basis of this study that the students used in this sample responded to the Study of Values test when taken in the conventional manner, in the way they actually believed rather than as they thought they "should believe." At least the conformity to social expectancy is different when students take the test in the conventional manner than when they are asked to conform to some sort of social expectancy.

### Summary

Ninety-three freshman women were given the *Allport-Vernon Study of Values* test three times, twice in the usual way and once with the students instructed to fill out the test as they felt they "should believe." Some statistically significant differences in value scores indicated that the students responded differently when instructed to take the test as they thought they "should believe" than when they took the test in the conventional manner.

### REFERENCES

1. Allport, G. W., and Vernon, P. E., *A Study of Values*. Boston: Houghton-Mifflin Company, 1931.
2. Cantril, H., and Allport, G. W. "Recent Applications of the Study of Values." *Journal of Abnormal and Social Psychology*, XXVIII (1933), 259-273.
3. Duffy, Elizabeth. "A Critical Review of Investigations Employing the Allport-Vernon Study of Values and Other Tests of Evaluative Attitude." *Psychological Bulletin*, XXXVII (1940), 597-612.
4. Fischer, Robert P. "Do the Interests of Students Indicate the Need of a Liberal Education?" *Journal of Educational Research*, XXXVII (1944), 619-627.
5. Spranger, E., *Lebensformen Geisteswissenschaftliches Psychologie und Ethik der Persönlichkeit*, 5. Aufl. Halle (Salle), M. Niemeyer, 1925, 450 pp.



## MEASUREMENT ABSTRACTS<sup>1</sup>

Altus, William D. "Some Correlates of Enuresis Among Illiterate Soldiers" *Journal of Consulting Psychology*, X (1946), 246-259.

A comparative study was made of 76 enuretic illiterate soldiers who were matched with a control group of non-enuretic illiterate soldiers on variables of chronological age, test intelligence, and race or linguistic grouping. Enuresis was found to be associated with racial and linguistic grouping, occupational instability, and incidence of failure in school, of arrests, and of venereal infection. An orally administered test of adjustment, validated for use at the Ninth Service Command Special Training Center, showed a perfectly reliable mean difference in score between the two groups, the most discriminating questions indicating for the enuretic group a syndrome represented by the descriptive psychiatric term "anxiety." *Frances Smith.*

---

Altus, William D. "The Validity of the Terman Vocabulary for Army Illiterates." *Journal of Consulting Psychology*, X (1946), 268-276.

The use of the Terman Vocabulary with illiterate army inductees in a special training center was found to yield valid results when an individual's score was interpreted in terms of norms for his own group. The criterion for validation was the graduation or discharge of the trainee from the center. The groups were distinguished as White, Colored, or bilingual, with four language groups differentiated in the latter. It was found that no words in the test retain equal validity for differing racial and bilingual groups, and that only White and possibly Colored distributions on the test approach normality. Fairly close agreement was found to exist between the *Wechsler Mental Ability Scale*, Form B, and the Terman Vocabulary in relative difficulty of these measures for the various groups studied. *Frances Smith.*

---

Barrett, Dorothy M. "Prediction of Achievement in Typewriting and Stenography in a Liberal Arts College." *Journal of Applied Psychology*, XXX (1946), 624-630.

For purposes of determining success in typewriting and stenography, 96 students of Hunter College taking courses in typing and 75 studying shorthand were given, prior to the beginning of classes, the following tests: the *Bennett Stenographic Aptitude Test*, the *Kuder*

---

<sup>1</sup> Edited by Forrest A. Kingsbury.

*Preference Record*, the *MacQuarrie Test for Mechanical Ability*, the *Minnesota Vocational Test for Clerical Workers*, the *Strong Vocational Interest Blank for Women*, the *Thurstone Vocational Interest Schedule*, and the *Tuise Shorthand Aptitude Test*. Of these, certain parts of the Minnesota, the MacQuarrie, and the Tuise, when used in combination, proved effective in differentiating between good and bad typists as well as yielding satisfactory predictions for success in both typewriting and stenography. *Fernon S. Truitt*.

---

Bettelheim, Bruno. "Self-Interpretation of Fantasy. The Thematic Apperception Test as an Educational and Therapeutic Device." *American Journal of Orthopsychiatry*, XVII (1947), 80-100.

The method of self-interpretation is discussed in connection with the diagnostic use of the T.A.T., case material being presented from a study of 32 female college students examined by their instructor. Such a method is found to be of value to the Examiner in making available additional associative material and in furnishing clues with regard to the character of the defenses and the depth of repression. It is of educational value in promoting general psychological insight on the part of the subject, and has therapeutic merit in helping him to understand the nature and origin of some of his own personality problems. Further investigation is needed on the part played in the usefulness of this method by the transference relationship existing between examiner and subject. *Frances Smith*.

---

Brogden, Hubert F. "Variation in Test Validity with Variation in the Distribution of Item Difficulties, Number of Items, and Degree of Their Intercorrelation." *Psychometrika*, XI (1946), 197-214.

The relation between item difficulty distributions and the "validity" and reliability of tests is computed through the use of normal correlation surfaces for varying numbers of items and varying degrees of item intercorrelations. Optimal or near optimal item difficulty distributions are thus identified for various possible item difficulty distributions. The results indicate that if a test is of conventional length, is homogeneous as to content, and has a symmetrical distribution of item difficulties, correlation with a normally distributed perfect measure of the attribute common to the items does not vary appreciably with variation in the item difficulty distribution. Greater variation was evident in the correlation with a second duplicate test (reliability). The general implications of these findings and their particular significance for evaluating techniques aimed at increasing reliability are considered. (Courtesy *Psychometrika*.)

---

Cattell, Raymond B. "Simple Structure in Relation to Some Alternative Factorizations of the Personality Sphere." *Journal of General Psychology*, XXXV (1946), 225-238.

Three alternative approximations to simple structure in rotating



the same set of variables from the personality sphere, rated on a population of 208 subjects, are examined. The last two structures lose five of the factors found in the original. However, the remaining factors appear to be substantially the same psychologically in all three results. As an additional aid to be used in conjunction with graphic plots in judging the quality of simple structure, the ratio of the frequency of projections from zero to half the mean projection to the frequency of projections from half the mean to the mean is suggested. This ratio would be found for each factor or as a mean value for all factors. The size of the dispersion of this ratio among the factors is also indicative. *Frederick Gehlmann.*

---

Clark, Walter H. "Perseverance and Repetition as Factors in Gain in IQ" *Journal of Educational Psychology*, XXXVII (1946), 557-562.

Boys attending a preparatory school over a period of 8-10 years were ranked by teachers on perseverance as exhibited in academic work. Yearly intelligence test scores, obtained from Otis Self-Administering, Ohio, California, and A C E., with A C E. scores converted to Otis units, were examined for changes in IQ, and the changes averaged. When average IQ changes in two extreme groups of persevering and non-persevering boys were compared, the difference between the two groups was found to yield a CR of only 1.45, and little correspondence was found between perseverance and gain in IQ within the contrasted groups. The increases in IQ occurring from year to year appeared, on comparison of results obtained from different tests and from different forms of the same test, to be due in significant degree to repetition of different forms of the same test. *Frances Smith.*

---

Davis, Frederick B. "The Factorial Composition of Two Tests of Comprehension in Reading" *Journal of Educational Psychology*, XXXVII (1946), 481-486

The author in a previous factorial study obtained two new measures of reading ability: Word-Knowledge and Reasoning in Reading. Tests for these two components were combined with a battery of fourteen Judgment-and-Reasoning tests and were administered to eleventh- and twelfth-grade boys in New York City. The speed of reading had no influence on scores. The answer sheets of 689 boys were analyzed. Intercorrelations of the 14 Judgment-and-Reasoning tests were computed and a factorial analysis of the resulting matrix was made, using the principal-axis method. Eight statistically significant components were obtained, two of which were similar to the Word-Knowledge and Reasoning-in-Reading components. Correlations between these two components and the 14 tests in the Judgment-and-Reasoning battery were computed. One of the principal conclusions of the author is that the type of reasoning ability in reading is different from that demanded by conventional reasoning tests. *Leroy Burwen.*

Des Lauriers, Austin and Halpern, Florence. "Psychological Tests in Childhood Schizophrenia." *American Journal of Orthopsychiatry*, XVII (1947), 57-67.

To throw some light on the psychological processes of the childhood schizophrenic, the authors administered a battery of tests to patients both before and after shock treatment. Test results reveal that disturbance in intellectual functioning is all-pervasive, affecting many areas, and that such patients demonstrate breaks in their successes and failures which are unrelated to the difficulty of the test items. The schizophrenic child encounters difficulty in organizing both the outside and inside world and in developing meaningful relationships. As long as he is beset by anxiety, he remains in some contact with reality. Post-shock results demonstrate a greater ability to concentrate, a flattening of effect, an absence of "fear of fear," and a reduction of fantasy life. *Harold Mowbray*.

Dolger, Laura and Ginandes, Janet. "Children's Attitudes Toward Discipline as Related to Socioeconomic Status." *Journal of Experimental Education*, XV (1946), 161-165.

Random samples from two New York City schools which differed markedly in educational philosophies and in the economic and social status of the students were used as criterion groups. Data were procured by two methods: The first was a written composition following the reading of a story involving two siblings, one of whom was a disciplinary problem; following this, a routine interview was given in which a series of ten questions relating to disciplinary factors were presented. The children in the school representing a low socio-economic level offered fewer constructive solutions to the disciplinary problem, tended more toward placing the responsibility upon the errant child rather than the provoking situation, and resorted more often to punishment by authorities, than did the children from the school with the higher socio-economic level. *Rae Shiffrin*.

Goodman, Charles H. "The MacQuarrie Test for Mechanical Ability: I. Selecting Radio Assembly Operators." *Journal of Applied Psychology*, XXX (1946), 586-595.

This describes the first of 4 experimental studies based on the *MacQuarrie Mechanical Ability Test*, the object being to determine its usefulness as a means of selecting personnel in radio manufacturing. The subjects were 329 female radio assembly operators, of age range 16 to 64, with a mean of 27.3 years, hired over a five-month period between November, 1943, and March, 1944. All were given the MacQuarrie test before the three-day training period, at the conclusion of which they were given a manual test to measure their mastery of the instruction. The adequacy of the MacQuarrie was determined by calculating Pearson's *r*'s of subjects' total test score and sub-test scores with the criterion which was the rating system of the Vestibule Training School, indicating it to be more efficient than the company's hiring methods. *Vernon S. Tracht*.

Gough, Harrison G. "The Relationship of Socio-Economic Status to Personality Inventory and Achievement Test Scores" *Journal of Educational Psychology*, XXXVII (1946), 527-540

It is suggested that the concept of socio-economic status is erroneously treated as a disparate unit and that status should be conceived of as a prestige variable dependent on social and economic factors not configured in any constant manner. Various studies of the correlation of social and economic status with measures of intelligence are reviewed, prior to the report of a study on 127 sixth-grade school children which was conducted to determine the effects of socio-economic status upon academic achievement and upon personality inventory scores. Intercorrelations between scores on the *American Home Scale*, the *Brown Personality Inventory for Children*, and various achievement tests show that in the sample studied, socio-economic status has a slight positive relation to academic achievement, while personality inventory scores have a slight negative relation to achievement and to status. *Frances Smith.*

---

Heim, A. W. "An Attempt to Test High-Grade Intelligence." *British Journal of Psychology*, XXXVII (1947), 70-81

The purpose of this investigation was to determine the possibility and feasibility of discriminating between individuals in the highest 10 per cent of the intelligence range. After discussing some of the limitations of existing tests, the author explains the design of her test, AH5, which includes verbal, arithmetic, and visual materials. She detects some degree of relationship between the rapidity of performance and the test score and a tendency for mathematicians and engineers to score higher than language students. No conclusive evidence is available for demonstrating the ability of AH5 to discriminate in the high intellectual levels. *Harold Morak*

---

Johnson, Donald M. "Technique for Analysis of a Highly Generalized Response Pattern." *Psychological Review*, LIII (1946), 348-361.

As an initial attack on the problem of analyzing patterned behavior, the responses of college students to number series were studied. The theory used for predicting behavior is based on the application of the generalization hypothesis. The generalization gradient fits three-number sequences very well. For four-number and five-number sequences it is necessary to combine the generalization principle and either an end-effect or a cyclical trend. These findings, based on 4590 cases, have some practical bearing on the interpretation of rating-scale results. *Frederick Gehlmann.*

---

Mintz, Alexander. "Reading Reversals and Lateral Preferences in a Group of Intellectually Subnormal Boys" *Journal of Educational Psychology*, XXXVII (1946), 487-501.

Following a discussion of Orton's neurological explanation of

reading disabilities, the author describes an experiment with 95 intellectually subnormal boys to test the theory of "strephosymbolia." He duplicates the findings of Orton, yet finds no evidence of "strephosymbolia." Emotional problems, for example, might conceivably explain the apparent relationship between weak lateral cerebral dominance and reading difficulties. He concludes by offering many possible explanations for reading disability without employing Orton's theory. *Harold Mosak.*

Pasmanick, Benjamin. "A Comparative Study of the Behavioral Development of Negro Infants." *Journal of Genetic Psychology*, LXIX (1946), 3-44.

Fifty-three Negro infants of New Haven, Connecticut, living with their families, were compared in behavioral development with three groups of white infants, one comprised of children living in boarding homes, another of those living in child-care institutions, and a third group of those from superior families living at home. No outstanding differences were found between the two races, except a pattern of acceleration in gross motor behavior displayed by the Negro infants. Negro infants showed "more equanimity and restraint" in the examination process. There were no pronounced differences in personality structure. The author attributes the equivalent developmental progress to the better maternal prenatal diets of the Negro population during the prosperous war years when these infants were born. *Rae Skjrin.*

Patterson, C. H. "A Comparison of Various 'Short Forms' of the Wechsler-Bellevue Scale." *Journal of Consulting Psychology*, X (1946), 260-267.

There is a need in clinical psychology for more good comprehensive tests concerned with diagnosis and understanding. When a brief form is called for by particular circumstances, it should be derived from a longer test that is available for more detailed examination, and should not only meet minimum time requirements but yield high correlation with full-scale score and be diagnostically useful. Five previously proposed "short forms" of the *Wechsler-Bellevue Scale* are compared in a single sample of 50 psychiatric patients; and also analyzed are two new short forms composed respectively of (1) Vocabulary, Comprehension, and Digit Symbol, and (2) Vocabulary, Comprehension, Block Designs, and Picture Completion. These two new forms are judged to meet best, in general, the criteria proposed, though the particular form utilized should depend on the situation and need. *Frances Smith.*

Peters, Henry N. "The Mirror-Tracing Test as a Measure of Social Adaptation." *Journal of Abnormal and Social Psychology*, XLI (1946), 437-448.

In extending the investigations performed on mirror tracing and its relationship to emotionality, the author presents both qualitative

and quantitative data for mirror tracing in socially adapted and maladapted groups. He discovers that mirror tracing is useful in eliciting the emotional reactions of his subjects. Furthermore, a significant difference in average time scores exists between adapted and maladapted groups. This leads the author to a "partially verified" hypothesis that "Subjects vary in the abstractness of their attitude in approaching the mirror tracing problem, while the problem is especially amenable to a concrete attitude" *Harold Mosak.*

---

Robinson, Mary Frances "What Price Lobotomy?" *Journal of Abnormal and Social Psychology*, XLI (1946), 421-436

Ten schizophrenic patients with prefrontal lobotomies were compared with 7 similarly diagnosed patients, who had not undergone this brain operation, to determine whether such injury to the prefrontal area results in unfavorable intellectual changes heretofore overlooked by psychiatrists in their enthusiasm regarding improvement in behavior. Such tests as the Shipley-Hartford, the Hunt-Minnesota, the Porteus Mazes, a measure of deliberation devised from other tests and standardized by the author, and 3 parts of the Downey Will-Temperament tests were included. Although exhibiting little or no difference from the control group of schizophrenics on the first 3 of these tests, the lobotomized individuals did significantly less well on the latter 2, i.e., those measuring degree of deliberativeness. Results of this study indicate that bilateral prefrontal lobotomy produces a definite mental deficiency—a reduction in the capacity for prolonged attention—irrespective of the therapeutic benefits. *Vernon S. Tracht*

---

Rosenzweig, Saul and Isham, A. C. "Complementary Thematic Apperception Test Patterns in Close Kin." *American Journal of Orthopsychiatry*, XVII (1947), 129-142.

The hypothesis is presented that the use of projective techniques with mental patients and their close relatives may reveal important complementary psychodynamics. The T.A.T. was employed in the study under consideration as being more readily adaptable than the Rorschach to correlative and dynamic use with patients and close kin. An illustrative case from the study is presented in detail, and the emerging complementary pattern is analyzed. It is pointed out that the method has advantages both in time saving and in certainty of interpretation over usual interview procedure, and that it is promising as a supplementary method of inquiry in areas of social and anthropological research as well as of psychopathology. *Frances Smith.*

---

Sanford, R. Nevitt, Conrad, Herbert S., and Franck, Kate. "Psychological Determinants of Optimism Regarding Consequences of the War." *Journal of Psychology*, XXII (1946), 207-235.

To determine the personality-correlates of optimism, 84 men and 119 women students of the University of California (September,

1942) were given a specially prepared scale of 14 items measuring optimism concerning war consequences, and a specially devised personality questionnaire. Generally speaking, optimism and a healthy personality structure were found to "go together," war consequences-optimism being related to numerous diverse factors covering very different areas of personality and social background. The sample of women yielded more significant, clear-cut results than that of the men. *Feynon S. Tracht*.

---

Simrall, Dorothy. "Intelligence and the Ability to Learn." *Journal of Psychology*, XXIII (1947), 27-43.

Since the definition of intelligence as the ability to learn has gained such wide acceptance, it is the responsibility of the psychologist to verify it. The author sets up nine hypotheses deduced from this theory of the nature of intelligence and employs the Thurstone method of factor analysis to test them. She finds consistent negative results in testing these hypotheses indicating that intelligence and the ability to learn may be independent variables. *Harold Mosak*.

---

Stone, Calvin P., Girdner, John, and Allmarch, Ruth. "An Alternate Form of the Wechsler Memory Scale." *Journal of Psychology*, XXII (1946), 199-206.

A second form of the *Wechsler Memory Scale* is made available by the authors to fulfil the need of those experimenters who require at least two comparable forms of a test to appraise variability of memory during and after electric shock therapy. The subtests of the new scale are in general identical with those of the first scale either "in fact or in principle." *Harold Mosak*.

---

Van Bruggen, John Andrew. "Factors Affecting Regularity of the Flow of Words During Written Composition." *Journal of Experimental Education*, XV (1946), 133-155.

The problem of the investigation was to determine how the rate of flow of words during the writing of compositions is affected by various compositional, academic, personal, and environmental factors. Eighty-four pupils from grades seven through nine of the Oakdale Christian School in Grand Rapids, Michigan, were asked to write two compositions, one original and one a reproduction of a story. The rate of flow was recorded by a kymograph. Reproducing from memory was found to require less time than original composition. Relations were established between the flow of words and scores in reading, vocabulary, spelling, and marks in English. The rate of flow increases with the chronological and mental age. Those who were rated as dominating, extroversive and persevering had a correspondingly swifter rate of flow. A positive relationship was found between writing by thought rather than by isolated words, and also with the quality of the composition. *Rae Shifrin*.

Verniaud, Willie Maude. "Occupational Differences in the Minnesota Multiphasic Personality Inventory." *Journal of Applied Psychology*, XXX (1946), 604-613.

Ninety-seven female subjects representing 3 contrasting occupations—40 clerical workers, 27 department store saleswomen, and 30 optical workers—were given the *Minnesota Multiphasic Personality Inventory* as a means of finding out whether job differences would be revealed on this Inventory. While the author cautions that the results of this study must be regarded in the light of specific job settings (i.e., they may not necessarily apply to workers doing similar tasks under very different job surroundings), she concludes that group differences exist in the personality of successful workers corresponding to gross differences in occupational requirements, these differences in turn being identified by responses on the MMPI. *Vernon S. Tracht.*

---

Verville, Elinor. "The Effect of Emotional and Motivational Sets on the Perception of Incomplete Pictures." *Journal of Genetic Psychology*, LXIX (1946), 133-145.

Nine incomplete pictures of common objects were presented by the projection method to six groups of 25 women students. One group was control. In each of the experimental groups, certain tasks were performed for 30 minutes prior to the presentation of the pictures in an attempt to produce certain emotional sets. The five sets were tension, complete failure and frustration, failure according to false norms, success according to false norms, and reaction to personality testing. Mean reaction times for each picture and for the total set indicate that the control, success, and failure by norms groups reacted more quickly than the complete failure and personality groups. *Frederick Gehlmann.*

---

Wallin, J. E. W. "A Comparison of the Stanford-Binet 1916 and 1937 (Form L) Test Results with Those from the Arthur Performance Scale (Form I) Based on the Same Subjects." *Journal of Genetic Psychology*, LXIX (1946), 45-55.

Results obtained from (1) 290 boys and girls on the 1916 *Stanford-Binet* (Form L) and the *Arthur Performance Scale* (Form I), and (2) 172 boys and girls on the 1937 *Stanford-Binet* (Form L) and the *Arthur Performance Scale* (Form I) are compared. A correlation of .72 was obtained for (1) and a correlation of .53 for (2). The average difference in intelligence age is less than 13 months for 60.8% of the cases on the 1916, and 48.8% on the 1937 *Stanford-Binet*. The Mean 1916 *Stanford-Binet* IQ is 3 points lower than the corresponding Arthur; it is 8 points for the 1937 scale. Possible explanations for these results are proposed by the author. *Leroy Burwen.*

Wherry, Robert J. "Test Selection and Suppressor Variables." *Psychometrika*, XI (1946), 239-247.

A theoretical discussion of the factor pattern of predictor tests and criterion shows that ordinary test selection methods break down under certain circumstances. It is shown that maximal results may not occur if suppressor variables are present among the predictors. Suggested solutions to the problem include: (1) prior item analysis of tests against the criterion, (2) selection of several trial batteries including some with suppressor variables on the basis of a factor analysis of tests and criterion, (3) modification of the usual test selection procedures to include separate solutions based upon each of several starting variables, or (4) the cumbersome and tedious solution of all possible combinations of predictors. The solutions are recommended in the order named above. Although all of the suggested solutions involve added labor and may not be necessary, the test or battery constructor should at least be aware of the problem. (Courtesy *Psychometrika*).

Winthrop, Henry. "Semantic Factors in the Measurement of Personality Integration." *Journal of Social Psychology*, XXIV (1946), 149-175.

The theory is advanced that degree of personality integration may be operationally defined by an index of attitude consistency. The author describes construction, method of presentation, and scoring of a test of attitude consistency composed of 100 pairs of attitude sentences logically devised on the basis of the Aristotelian Diagram of Opposition to represent either contradictory or contrary or equivalent forms of attitudes toward identical subject matter. Experimental findings are presented for a college population of 253, with 31 subgroups. It is suggested that the most important causes underlying attitude-inconsistency are those generated in semantic blockage, human conflict in behavior being due in substantial proportion to improper manipulation of verbal aspects of the environment. *Frances Smith*.

(Film). *Examining Personnel for Civilian Employment*. Research Section, Civilian Personnel Division, Office, Secretary of War. Four film strips with accompanying phonograph records. 15 minutes each. Available through loan. Washington 25, D. C.

This film presents the problem of selecting good workers from among many applicants who are available. It describes the pattern of procedures developed for the employment of civilian personnel for a wide range of positions in War Department installations throughout the country. The first step in the production and use of selection tests is the appointment of local examining boards, composed of operating officials in various occupational fields. Examining committees or panels are then appointed to prepare tests for each type of job. These panels begin their work by making an analysis of the duties of the jobs to be filled. Sources of information for job



analyses and steps in the work are illustrated. The first three film strips and narration deal with the preparation of an actual examination for a particular job, and describe in detail the job analysis, the determination of essential personal characteristics for doing the work, and the development, administration, and validation of the examination. The fourth film strip and narration describe in general terms the use of such examining techniques as the biographical information blank, oral examination, standard interview, written tests, performance tests, and reference schedules. It deals with both proficiency and aptitude tests and shows their applicability to different types of jobs. *Charles I. Mosier.*

#### ADDITIONAL ARTICLES NOT ABSTRACTED

- Alper, Thelma G. "Memory for Completed and Incompleted Tasks as a Function of Personality: an Analysis of Group Data." *Journal of Abnormal and Social Psychology*, XLI (1946), 403-420.
- Beckham, Albert. "Albinism in Negro Children." *Journal of Genetic Psychology*, LXIX (1946), 199-215.
- Bolgar, Hedda and Fischer, Liselotte. "Personality Projection in the World Test." *American Journal of Orthopsychiatry*, XVII (1947), 117-128.
- Brodbeck, Arthur J. and Irwin, Orvis C. "The Speech Behavior of Infants Without Families." *Child Development*, XVII (1946), 145-156.
- Cronbach, Lee J. "A Case Study of the Split-Half Reliability Coefficient." *Journal of Educational Psychology*, XXXVII (1946), 473-480.
- Cummings, Jean D. "A Follow-up Study of Emotional Symptoms in School Children." *British Journal of Educational Psychology*, XVI (1946), 163-177.
- Davis, C. Jane. "Correlation Between Scores on Ortho-Rater Tests and Clinical Tests." *Journal of Applied Psychology*, XXX (1946), 596-603.
- Davis, Frederick B. "A Brief Comment on Thurstone's Note on a Reanalysis of Davis' Reading Tests." *Psychometrika*, XI (1946), 249-255.
- Dubin, Elizabeth Ruth. "The Effect of Training on the Tempo of Development of Graphic Representation in Preschool Children." *Journal of Experimental Education*, XV (1946), 166-173.
- Ellis, Albert. "Discussion of Heinlein's Comment on 'The Validity of Personality Questionnaires'." *Psychological Bulletin*, XLIV (1947), 83-86.
- Embree, R. B., Jr. "Note on the Estimation of College Aptitude Test Scores from IQ's Derived from Group Intelligence Tests." *Journal of Educational Psychology*, XXXVII (1946), 502-504.
- Engle, T. L. "The Use of the Harrower-Erickson Multiple Choice (Rorschach) Test in Differentiating Between Well Adjusted and

- Maladjusted High School Pupils." *Journal of Educational Psychology*, XXXVII (1946), 550-556.
- Gustav, Alice. "Estimation of Rorschach Scoring Categories by Means of an Objective Inventory." *Journal of Psychology*, XXII (1946), 253-264.
- Harris, Daniel H. "Questionnaire and Interview in Neuropsychiatric Screening." *Journal of Applied Psychology*, XXX (1946), 644-648.
- Heinlein, Paul. "Comment on 'The Validity of Personality Questionnaires.'" *Psychological Bulletin*, XLIV (1947), 80-82.
- Himmelfreit, H. T. "A Comparative Study of the Level of Aspiration of Normal and of Neurotic Persons." *British Journal of Psychology*, XXXVII (1947), 41-69.
- Horrocks, John E. and Thompson, George G. "A Study of the Friendship Fluctuations of Rural Boys and Girls." *Journal of Genetic Psychology*, LXIX (1946), 189-198.
- Howard, Ruth W. "The Language Development of a Group of Triplets." *Journal of Genetic Psychology*, LXIX (1946), 181-188.
- Irwin, Francis W. and Gebhard, Mildred E. "Studies in Object Preferences: The Effect of Ownership and Other Social Influences." *American Journal of Psychology*, LIX (1946), 633-651.
- Kerr, Willard A. "Worker Attitudes Toward Scheduling of Industrial Music." *Journal of Applied Psychology*, XXX (1946), 575-578.
- Klugman, Samuel. "Relationship Between Performance on the Rotter Aspiration Board and Various Types of Tests." *Journal of Psychology*, XXIII (1947), 51-54.
- Lefever, D. Welty, Van Boven, Alice, and Banater, Joseph. "Relationship of Job Information Test Scores to Other Personnel Data for Enlisted Air Corps Men." *Occupations*, XXV (1947), 220-221.
- Napoli, Peter J. "Finger-Painting and Personality Diagnosis." *Genetic Psychology Monographs*, XXXIV (1946), 133-230.
- Napoli, Peter J. "Interpretive Aspects of Finger-Painting." *Journal of Psychology*, XXIII (1947), 93-132.
- Nemec, Lois G. "Relationships Between Teacher Certification and Education in Wisconsin: A Study of Their Effects on Beginning Teachers." *Journal of Experimental Education*, XV (1946), 101-132.
- Patterson, C. H. "On the Problem of the Criterion in Prediction Studies." *Journal of Consulting Psychology*, X (1946), 277-280.
- Recktenwald, Lester N. "Homogeneity of Items in Cleeton's Classification." *Occupations*, XXV (1947), 275-277.
- Reynolds, William A. "A Prepunched Master Deck for the Computation of Square Roots on I B M Electrical Accounting Equipment." *Psychometrika*, XI (1946), 223-238.
- Rogers, R. C. "Analysis of Two Point-Rating Job Evaluation Plans." *Journal of Applied Psychology*, XXX (1946), 579-585.

- Rose, Annelies Argelander. "The Effect of the War on the Social and Emotional Adjustment of College Girls." *Journal of Social Psychology*, XXIV (1946), 177-185.
- Sartain, A. Q. "A Note on Otis S.-A. Test Scores Made by Candidates for Supervisory Positions in an Aircraft Factory." *Journal of Social Psychology*, XXIV (1946), 253-254.
- Sartain, A. Q. "A Study of Bernreuter Personality Inventory Scores Made by Candidates for Supervisory Positions in an Aircraft Factory." *Journal of Social Psychology*, XXIV (1946), 255-259.
- Schmidl, Fritz. "The Rorschach Test in Juvenile Delinquency Research." *American Journal of Orthopsychiatry*, XVII (1947), 151-160.
- Seago, Dorothy W. "Stereotypes: Before Pearl Harbor and After." *Journal of Psychology*, XXIII (1947), 55-63.
- Stoughton, M. Louise. "A Study of Children's Heroes and Ideals." *Journal of Experimental Education*, XV (1946), 156-160.
- Strang, Ruth. "Methods, Technics, and Instruments of Mental Hygiene Diagnosis and Therapy." *Review of Educational Research*, XVI (1946), 436-443.
- Stuit, Dewey B and Wilson, John T. "The Effect of an Increasingly Well Defined Criterion on the Prediction of Success at Naval Training School (Tactical Radar)." *Journal of Applied Psychology*, XXX (1946), 614-623.
- Thompson, George G. and Witryol, Sam L. "The Relationship Between Intelligence and Motor Learning Ability, as Measured by a High Relief Finger Maze." *Journal of Psychology*, XXII (1946), 237-246.
- Tinker, Miles A. and Paterson, Donald G. "Readability of Mixed Type Forms" *Journal of Applied Psychology*, XXX (1946), 631-637.
- Vernon, P. E. "An Experiment on the Value of the Film and Film-Strip in the Instruction of Adults." *British Journal of Educational Psychology*, XVI (1946), 149-159.
- Wall, W. D. "Reading Backwardness Among Men in the Army" *British Journal of Educational Psychology*, XVI (1946), 133-148.
- Weichelt, John A. "A First Order Method for Estimating Correlation Coefficients." *Psychometrika*, XI (1946), 215-221.
- Welch, Livingston. "Recombination of Ideas in Creative Thinking." *Journal of Applied Psychology*, XXX (1946), 638-643.
- Wells, F. L. "Verbal Excess Over Quantitation: Two Case Studies." *Journal of Psychology*, XXIII (1947), 65-82.
- Wells, F. L. "Verbal Facility: Positive and Negative Associations." *Journal of Psychology*, XXIII (1947), 3-14.
- Wesman, Alexander G. and Bennett, George K. "The Use of 'None of These' as an Option in Test Construction." *Journal of Educational Psychology*, XXXVII (1946), 541-549.
- Wheatley, Luis Andres and Sumner, F. C. "Measurement of Neurotic Tendency in Negro Students of Music." *Journal of Psychology*, XXII (1946), 247-252.



## THE CONTRIBUTORS

Avonne L. Andrews—B.S.C., University of Illinois, 1944. Research Assistant, Division of General Studies, University of Illinois, 1944-1946.

William Brody—M P A, New York University, 1943. Member of examining staff, New York City Civil Service Commission, 1937-1942. Assistant Regional Personnel Officer, Office for Emergency Management, 1942-1943. Chief of Personnel, U S. Treasury Procurement Division, 1943-1944. Director of Personnel, National War Labor Board, 1944-1945. Director of Personnel, National Wage Stabilization Board, 1946. Director of Personnel, New York City Department of Health, 1945-. Member, American Public Health Association, American Society for Public Administration, Civil Service Assembly, Society for Personnel Administration.

Raymond B. Cattell—Ph.D., University of London, 1929; D Sc. (*ibid.*), 1939. Director of City Child Guidance Clinic, Leicester, England, 1932-1937. Research Associate to Professor Thorndike, Teachers College, Columbia University, 1937-1938. G. Stanley Hall Professor, Clark University, 1939-1941. Lecturer, Harvard University and Civilian Consultant, Adjutant General's Office, 1941-1944. Research Professor in Psychology, University of Illinois, 1945-. Author of *A Guide to Mental Testing*, *Crooked Personalities in Childhood and After*, *General Psychology*, *The Description and Measurement of Personality*, and other books on personality and social psychology, as well as of research articles in American and British journals. Member, American Psychological Association. Fellow, British Psychological Society.

Robert P. Fischer—Ph.D., Ohio State University, 1942. Member of the Faculty at Clark University, 1942-1943. At the University of Illinois, 1943-Spring, 1946. At Birmingham-Southern University, Spring, 1946-Fall, 1946. At present, Associate Professor of Psychology, University of Florida.

Wiley S. Garrett—M.A., West Virginia University, 1933. School Psychologist, Warren City Schools, Warren, Ohio, 1933-1943. Director of Special Education, Niles City Schools, Niles, Ohio, 1943-. Author of articles in *Ohio Schools* and *Occupations*. Associate Member, American Psychological Association.

Louis Guttman—Ph.D., University of Minnesota, 1942. Teaching Assistant in Sociology, 1936-1940; Acting Director, Sociological Research Laboratory, 1939-1940, University of Minnesota. Pre-doctoral Field Fellow, Social Science Research Council, 1940-1941. Instructor in Sociology, Cornell University, 1941-1942. Expert Consultant to the Secretary of War, Research Branch, Information and Education Division, War Department, 1941-1946. Assistant Professor, Associate Professor of Sociology, Cornell University, 1942-. Co-author of *Prediction of Personal Adjustment*. Author of articles on attitude and public opinion analysis, scale and intensity analysis, factor analysis, test-re-test reliability, and other techniques for the analysis of attitudes, surveys, and achievement. Collaboration Editor, *Journal of the American Statistical Association*, 1945-. Fellow, Institute of Mathematical Statistics, American Association for the Advancement of Science. Member, American Sociological Society, American Statistical Association, Psychometric Society, Sigma Xi.

Lindsey R. Harmon—M.S., 1938, and two years additional graduate work, University of Minnesota. Psychologist, American Red Cross Department for the Handicapped, St. Louis, 1937-1941. Research Assistant, Teaching Assistant, Instructor, University of Minnesota, 1941-1943. Vocational Adviser and Vocational Advice-ment Supervisor, Veterans Administration, 1943-. Member, American Psychological Association, Psi Chi, Minneapolis Vocational Guidance Association. Executive Secretary, Minnesota Society for Applied Psychology, 1944-.

Chester W. Harris—Ph.D., University of Chicago, 1946. Teacher, Public Schools, Denver, Colorado, 1935-1939. Research Assistant, Evaluation Staff of the Eight-Year Study, University of Chicago, 1939-1941. Aviation Psychologist, Army Air Forces Aviation Psychology Program, 1942-1946. Assistant Professor of Education and Examiner, University of Chicago, 1946-. Member, Kappa Delta Pi, Phi Delta Kappa. Associate, American Psychological Association.

Joseph C. Heston—Ph.D., Ohio State University, 1941. Science Instructor and Director of Remedial Work, West Jefferson, Ohio, High School, 1932-1939. Assistant in Psychology, Ohio State University, 1939-1941. Instructor in Psychology, 1941-1942; Assistant Professor of Psychology, 1942-1946; Director of Bureau of Testing and Research, 1944-; Associate Professor of Psychology, 1946-, De Pauw University. Research Consultant, Farm Security Administration, 1940-1942. Author of articles in *Educational Research Bulletin*, *Journal of Educational Psychology*, and *Journal of Applied Psychology*. Member, American Psychological Association, American College Personnel Association, American Association of University Professors, Sigma Xi.

Charles I. Mosier—Ph D., University of Chicago, 1937. Instructor in Psychology and Vocational Guidance Counselor, University of Florida, 1933–1936. Assistant Professor of Psychology, University of Florida, 1937–1939. Acting University Examiner, University of Florida, 1938. Assistant Examiner, Sloan Research Project, 1940–1941. Personnel Research Technician, State Technical Advisory Service, Social Security Board, 1941; Chief of Position Classification, 1942; Chief of Personnel Methods and Standards, 1943–1944; Chief of Research and Test Construction, 1945–1946. Office of the Secretary of War, Acting Chief, Research Section, 1946–. Author of articles in *Psychometrika*, *Psychological Review*, *Journal of Educational Psychology*, and other journals. Associate Member, American Psychological Association. Member, Psychometric Society, Southern Regional Committee of the Social Science Research Council. Member of the editorial boards of *Psychometrika* and *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*.

William A. Owens—Ph.D., University of Minnesota, 1940. Instructor in Psychology, Iowa State College, 1940–1942. On leave for service with the N. S. Navy, Research Activity, Bureau of Naval Personnel, Washington, D. C., 1942–1945. Associate Professor and Chief Appraiser V.A.G.C., 1945–1946, Professor of Psychology, 1947–, Iowa State College. Author of articles on differential psychology and measurement. Associate Member, American Psychological Association. Member, American Association for the Advancement of Science, Iowa Academy of Science.

Norman J. Powell—Ph.D., Columbia University, 1936. Psychologist, Psychological Corporation, 1930–1931. Psychologist, Sing Sing Prison, 1931–1936. Personnel Examiner, Director of Research, Director of Classification and Research, New York City Civil Service Commission, 1936–1941. Lecturer, Brooklyn College, 1939–1942. Instructor, Hunter College, 1941. Consultant, Rhode Island Civil Service Department, 1939–1940. Consultant, New Hampshire Position and Compensation Survey, 1942–1943. Consultant, New York State Legislature, 1944–1945. Personnel Consultant, Philadelphia Council of Social Agencies, 1944. Personnel Consultant, Office of Price Administration, 1945. Instructor, College of the City of New York, 1942–. Author of articles on personnel administration and public administration. Member, American Political Science Association, American Society for Public Administration, Municipal Personnel Society of New York, American Association of University Professors.

(Mrs.) Jean Waid Reilly—M.A., Ohio State University, 1942. Assistant Personnel Adviser to Women, 1942–1943; Personnel Adviser to Women, 1943–1944, University of North Carolina. Personnel Counselor, Rehabilitation Program, U. S. Naval Hospital, Pensacola, Florida, 1944–1945. Assistant, Student Psychological

Consultation Service, Ohio State University, spring 1946. Teaching Assistant, Department of Psychology, Ohio State University, 1946-.

Francis P. Robinson: Ph.D., University of Iowa, 1932. Instructor, University of Iowa, 1932-1933. Chairman, Department of Psychology, University of Iowa, 1933-1937. Assistant Professor of Psychology, University of Oregon, 1937. Assistant Professor of Psychology, 1937-1941; Associate Professor of Psychology, 1941-1944; Professor of Psychology, 1944-, Ohio State University. Author of *Diagnostic and Remedial Techniques for Effective Study, Psychology and the New Education*; co-author of *Effective Study*, and author of articles on educational psychology and student personnel. Fellow, American Psychological Association.



EDUCATIONAL and  
PSYCHOLOGICAL



MEASUREMENT

**VOLUME SEVEN, NUMBER THREE, AUTUMN**

<i>Suggestions for Writing Achievement Exercises to be Used in Tests Scored on the Electric Scoring Machine.</i> MAX D. ENGELHART .....	357
<i>Vocational Interests and Vocational Choice: Present Knowledge and Future Research in Their Relationships.</i> DONALD E. SUPER .....	375
<i>The Selection of Foremen.</i> MILTON M. MANDELL .....	385
<i>Measuring the Teaching Personality.</i> WALTER W. COOK and CARROLL H. LEEDS .....	399
<i>The Insight of College Students into the Items on a Personality Test.</i> J. A. MORRIS KIMBER .....	411
<i>A Study in Tool Usage.</i> DAVID L. EKBERG .....	421
<i>A Technique for Obtaining Summary Data from Aggregate Weighting Sheets on the Graphic Item Counter Attachment to the Test Scoring Machine.</i> CLIFFORD P. FROELICH and ROBERT J. KELLER .....	429
<i>Table for Use in the Computation of Statistics of Dichotomous and Truncated Distributions.</i> E. K. TAYLOR and R. H. GAYLORD .....	441
<i>Measurement Abstracts</i> .....	457
<i>The Contributors</i> .....	467

PRINTED IN THE UNITED STATES OF AMERICA  
THE SCIENCE PRESS PRINTING COMPANY  
LANCASTER, PENNSYLVANIA

# SUGGESTIONS FOR WRITING ACHIEVEMENT EXERCISES TO BE USED IN TESTS SCORED ON THE ELECTRIC SCORING MACHINE

MAX D. ENGELHART  
Chicago City Junior College

A NUMBER of colleges or universities and city school systems have established departments in which International Business Machines Corporation electric scoring machines are used to score tests. Frequently the kinds of tests scored in this way are restricted to standardized psychological or achievement tests. In a growing number of higher institutions and school systems, however, locally constructed achievement tests are produced for machine scoring. Sometimes these tests are produced by full-time examiners. More often the tests are prepared by teachers working in cooperation with members of an examination staff. This is the case in the Chicago City Junior College where such efforts are directed by the author of this paper.

The production of locally constructed achievement tests can be greatly facilitated by making available to the participating teachers mimeographed directions for the writing of test exercises. Unless the instructions are quite explicit, teachers tend to produce exercises which are not well adapted to machine scoring, or which are otherwise faulty. Many teachers have difficulty in phrasing adequate directions for a series of exercises. Some teachers are likely to forget that the standard answer sheet limits the number of answers to a given exercise to no more than five. Few teachers realize that it is possible to use a variety of forms. Hence it is desirable to provide teachers with examples of various forms, along with the directions necessary for the writing of exercises of these types. The

rest of this article describes the kind of material which may be given to teachers as a means of directing and stimulating the production of achievement test exercises. It is the hope of the author that persons concerned with the development or improvement of achievement testing in higher institutions or in city school systems will feel free to reproduce this material in its original or modified form.<sup>1</sup>

*General Directions for Writing Achievement Exercises for Machine Scoring.*—The exercises used in tests to be scored on the electric scoring machine must conform to certain patterns so that the students will have no difficulty in recording their answers on the answer sheets. No exercise or test item can have more than five answers including correct and incorrect ones. Each exercise number on the answer sheet is followed by lettered spaces for only five answers. It is possible to write exercises having more than one answer of the five correct, but this practice is not recommended. Where certain series of exercises call for more than one correct answer to each exercise,

<sup>1</sup> The directions assume the use of the standard 150 or 300 exercise answer sheet with answer spaces lettered A, B, C, D, and E. (In the opinion of the author, the lettering of answer spaces is preferable to the numbering of them from 1 to 5, since it eliminates referring to two sets of numbers in test directions.) If hand scoring is contemplated and the pupils or students are to record their answers in the test booklets, the first part of the directions given for true-false or classification exercises will need to be phrased "On the blank which precedes each of the following statements, write the letter....." (Each statement or item will be preceded by a blank rather than a number.) Where an answer sheet is prepared for rapid hand scoring in which answer sheet numbers are followed by lettered squares, for example,

123. 

--	--	--	--	--

the phrase "blacken space" in the directions for true-false or classification exercises need merely be changed to "place a cross (X) in lettered square ....." In the case of series of matching or multiple-answer exercises the directions given later may be changed to "On the blank which precedes each of the following exercises, write the letter which designates.....," or "After the number on the answer sheet which corresponds to that of each of the following exercises, place a cross (X) in the one lettered square which designates....." Scoring stencils can be made easily by preparing a keyed answer sheet and punching holes in the spaces which represent the correct answers. An IBM punch is the most convenient device to use since it will reach anywhere into an 8½×11" page. The suggestions given above will prove useful where no scoring machine is available. The preparation of exercises for hand-scored answer sheets is an effective means of developing test materials which will later prove well adapted to machine scoring. A final suggestion or two may be made with respect to the hand-scored answer sheet. The letters A, B, C, D, and E need not appear above each group of five squares. It is sufficient that they appear at the top of each column. An attractive hand-scored answer sheet may be typewritten on good bond paper whose dimensions are 10×13", the squares may be drawn in India ink using a ruling pen, and the manuscript reproduced in quantity by planographing or lithoprinting. In printing, the sheet is reduced to 8½×11". Such a sheet may contain groups of squares numbered from 1 to 150.

students are stimulated to mark in the same way in other series of exercises where only one answer per exercise is expected. The practice of asking for more than one answer complicates scoring whether the score is simply to be the number right or the number right minus some fraction of the number wrong. While it is desirable to prepare exercises similar to the types described in later paragraphs, this does not mean that measurement is restricted to the memory of isolated facts. The content of achievement exercises has more influence than their form on the nature of the responses made by the students. Abilities other than mere memory are tested when the content is to some extent novel and when the selection of the correct answer requires discrimination. On the other hand, measurement of memory abilities is justified when the content relates to important facts, concepts, or principles.

In the production of objective exercises for machine scoring, serious thought should be given to the planning of the distribution of exercises of various types in order to secure both representative sampling of the subject matter of the course and of a variety of abilities. While it is possible to weight different series of exercises differently in machine scoring, it is much easier to secure appropriate weighting by having the numbers of items or exercises pertaining to each division of subject matter proportional to the importance of the subject matter division. If the general organization of the test as a whole is to follow that of the course, a variety of abilities may be tested within each series of exercises. For example, certain exercises of a series of multiple-answer exercises may require no more than the ability to remember the term defined by the introductory part of each exercise. Other exercises in the same series may require the functioning of abilities transcending memory, if the content of the exercises is to some extent novel. While the introduction of novel content is essential, if more than the memory of facts is to be tested, the novel content should be such that the student is able to determine the correct answer by means of thinking about facts he has had an opportunity to learn. This may be illustrated by an example from the field of physics. Suppose that the students have studied Archi-

medes' Principle that a body is buoyed up by a force equal to the weight of the fluid displaced. Suppose further that the applications of this principle have largely or exclusively been with respect to bodies floating or immersed in liquids. Presume also that the students have learned at some time during the course that both liquids and gases are fluids and that gases, particularly, become less dense as they become warmer. Then, the students have had the opportunity to acquire the facts needed in answering a novel and thought-provoking multiple-answer exercise based on the question: "Will a dirigible balloon go up more rapidly in warm air or in cold air?"

Discriminative thinking is promoted by presenting the student with plausible and somewhat related incorrect answers as well as by the use of novel content or by the presenting of more or less familiar content in unanticipated ways. Discriminative thinking is also promoted by preparing a series of related exercises. More is said about these matters in later paragraphs.

*True-False.*—If you must write true-false exercises, avoid writing obviously trivial or meaningless ones. Avoid making broad generalizations which are obviously true or false, i.e., avoid writing statements involving such terms as "always," "never," "none," "only," "all," and "every." Such terms are permissible, however, in intrinsically difficult statements, for example, "All persons born in the United States are citizens of the United States" and "All amphibia live in fresh water." Avoid items that are partly true and partly false, for example, "Cases in equity are sometimes tried by a judge and sometimes by a jury." Avoid items which express opinionated views, unless the purpose is to test the knowledge of the source of an opinion, as for example, "According to Keynes, government spending is an excellent means of combating depressions." In items designed to measure the knowledge of an important concept or principle, avoid unnecessary technical terms or obscure minutiae. Avoid the writing of unusually long and involved statements. Such statements are more often true than false and test-wise students realize that this is the case. The principles just mentioned also apply to other types of objective exercises. The following directions should precede each series of true-false statements in the completed test:

After the number on the answer sheet which corresponds to that of each of the following statements, blacken space

*A* if the statement is true,

*B* if the statement is false.

Be sure to randomize the answers. Avoid grouping more than two or three true statements or two or three false statements together. Random variation in the answers should also characterize the series of exercises of the types described below. Not more than three adjacent exercises or items should be answered by the marking of the same lettered space.

True-false exercises may be written on  $3 \times 5$  cards, one exercise to a card. The truth or falsity of the exercise should be indicated in red pencil. Writing exercises on separate cards facilitates the rejection of poor exercises. It is also the best way to proceed if it is the intent to accumulate a file of exercises for future use. If the exercises pertain to a particular chapter of a text or a unit of a syllabus, it is desirable to specify on the card the chapter or unit to which the exercise pertains in order to facilitate checking the accuracy of the phraseology of the exercise and the correctness of the key. The method also facilitates representative sampling of the content of the course. These remarks also apply to exercises other than true-false with the difference that larger cards, or even  $8\frac{1}{2} \times 11$  sheets of paper, may be required in writing them.

*Multiple-Choice.*—To reduce guessing, four or five answers should always be given. One of the answers should be definitely correct and the others *should be plausible* although incorrect. Avoid consistently writing correct answers which are longer than incorrect ones. The incorrect answers may be similar in form to the correct answers, opposite in meaning to the correct answers, or slightly less precise or complete than the correct answers. (In the latter cases students may be asked to mark the "best" answer.) The exercises should not be unusually long or complex. The answers may be single words or brief phrases, but, unless the exercise begins with a question, *each answer must complete the introductory sentence grammatically*. The exercises may be conveniently written on  $4 \times 6$  cards, one exercise to a card. The correct response

should be indicated by placing the letter of the correct response in red pencil to the left of the exercise and the source of the exercise should be given. An example follows:

- |   |   |   |
|---|---|---|
| C | <p>In the prosperous years of the '20's prices were high, money circulated rapidly, and credit expanded. During the first years of the '30's, a period of depression, prices were low, money circulated slowly, and credit contracted. This illustrates (A. a secular trend, B. alternative uses of capital, C. the quantity theory, D. factors of production, E. Gresham's law.)</p> | <p>Atteberry<br/>and others,<br/>pp. 473-474.</p> |
|---|---|---|

This example also illustrates the effective use of introductory material. It is best to include in the "item stem," or introductory part of the exercise, the phraseology that represents the problem of the exercise. For example, an exercise which begins, "The obstruction of legislation by endless speechmaking and other dilatory tactics is called . . ." and which concludes with such answers as "A. a caucus, B. filibustering, C. a gerrymander, D. logrolling, E. cloture." is superior to one that begins "A filibuster is . . ." and concludes with five long answers.

Where the answers to a multiple-choice exercise are longer than single words or very brief phrases, it is desirable to list the answers. Note that each answer of the first two exercises grammatically completes the sentence with which each of these exercises begins. In the following examples, note also the difference in style where the "item stem" is a question. Both methods of writing are equally desirable.

150. The fatigue of muscle is due primarily to
  - A. the overuse of the individual muscle fibers.
  - B. the production of lactic acid within the muscle cells.
  - C. excessive carbon dioxide production.
  - D. a limitation of the oxygen supply.
  - E. a limitation of the food supply.
172. The novel *Madame Bovary* begins and ends with Charles, because
  - A. it is not customary for a novelist to begin his story with the main character.
  - B. the plot demands it, since he is older than Emma and survives her.



- C. the author wishes us to see clearly what kind of a man Charles is.
  - D. Charles symbolizes the world in which Emma saw herself an imprisoned butterfly.
193. The United States Government under the Articles of Confederation was most successful in meeting which of the following problems?
- A. The raising of money to pay our debts to France and Holland.
  - B. The regulation of trade between the states.
  - C. The organization of the Northwest Territory.
  - D. The alleviation of social discontent.
  - E. The making of commercial treaties with foreign nations.

Where multiple-answer exercises are problems in mathematics, or are exercises with numerical answers as in chemistry and in physics, it is an effective device to use the phrase "None of the above answers" as answer *E*. Occasionally this phrase should represent the correct answer. It should not be used as an incorrect answer, however, if the correct answer, listed as an A, B, C, or D answer, is an approximation. In this case, a student could argue that the exercise has two correct answers.

The following directions should precede each series of multiple-choice exercises in the completed test:

After the number on the answer sheet which corresponds to that of each of the following exercises, blacken the *one* lettered space which designates the correct answer.

(If the exercises are of the "Best" answer type substitute the word *best* for *correct* in the directions above.)

*Matching.*—Each exercise may consist of three definitions to be matched with three terms included in a list of five. The definitions should concern terms that are likely to be confused by the student whose knowledge is not precise. Furthermore, the two extra terms should be good distracters. Similar exercises may pertain to content other than definitions and the terms may be names of places, personages, formulas, or other brief answers. The exercises should be written on 4 × 6 cards with the *key* and sources indicated in red. To promote discriminative thinking all of the items in each group of three should pertain to related concepts. Usually, the items of a

given group should be drawn from the same chapter or unit of subject matter. If these suggestions are followed, repeated use of the exercises is facilitated. Two examples from series of exercises in different subjects are given below:

- |   |                   |
|---|-------------------|
| 187. The process in which electrons are gained in the outermost orbit of the atoms of the element.  | A. Hydrogenation  |
| 188. The process in which the hydrogen and hydroxyl ions in solution unite to form water, and the other ions unite to form a salt.                                    | B. Ionization     |
|   | C. Neutralization |
| 189. The process in which electrons are lost from the outermost orbit of the atoms of the element.  | D. Oxidation      |
|   | E. Reduction      |
| <hr/>   |                   |
| 214. This victory ended an attempt to cut off New England from the rest of the colonies and was a major factor in the obtaining of the alliance with France.          | A. Bunker Hill    |
| 215. After Washington's retreat from New York, hope was renewed in the American cause by this victory.  | B. Yorktown       |
|   | C. Saratoga       |
| 216. This victory involved a feint toward New York followed by strategy which resulted in a large body of enemy troops being cut off from help by either land or sea. | D. Monmouth       |
|   | E. Trenton        |

Each group of three numbered items should be set off as indicated above. The entire series should be preceded by the following directions:

After the number on the answer sheet which corresponds to that of each of the following items, blacken the *one* lettered space which designates the term at the right to which the item correctly refers.

*Classification.*—There must be not more than five categories in an exercise, but the number of items may vary. The categories should be in some way related to each other, for if one category is unrelated to the others, the items which pertain to it are too easily identified. The numbered items may be phrases or complete sentences. In a given series all of the items should be similar in construction. Such series of items

may be written on  $8\frac{1}{2} \times 11$  sheets of paper. To facilitate repeated use, all of the items should pertain to the same chapter in a text, to related chapters, or in general to related subject matter. The sources of the items should be indicated on the same page as the items, usually in red to prevent a typist from including them in the finished test. The key should be indicated by writing the appropriate letter in red to the left of each item. Several examples of classification exercises are given below. In each case only a few samples of items are listed below each set of directions. *All of the types of categories are applicable to a variety of subject-matter fields.*

After each item number on the answer sheet, blacken space

*A* if the item is true of the Archeozoic Era.

*B* if the item is true of the Proterozoic Era.

*C* if the item is true of the Paleozoic Era.

*D* if the item is true of the Mesozoic Era.

*E* if the item is true of the Cenozoic Era.

- 
149. The era in which the Laurentian Revolution occurred. Life was probably restricted to primitive one-celled plants and animals.
  150. Four great glaciations occurred in the most recent period of this era.
  151. This era was characterized by the rise of flowering plants, modern insects, and primitive mammals and birds.
- Etc.

---

After each item number on the answer sheet, blacken space

*A* if the item is true of the Monroe Doctrine.

*B* if the item is true of the Open Door Policy.

*C* if the item is true of both the Monroe Doctrine and the Open Door Policy.

*D* if the item is true of neither the Monroe Doctrine nor the Open Door Policy.

- 
261. By adopting this policy the United States sought to safeguard important interests of the American people.
  262. According to this policy the interests of the United States take precedence over those of any European country.
  263. Violation of this policy occasioned the enunciation of the "Stimson Doctrine."
  264. Our traditional policy of freedom of the seas is basic to this policy.
- Etc.
-

After each item number on the answer sheet, blacken space  
*A* if the statement is true, and it is supported by the reason given.

*B* if the statement is true, but not because of the reason given.

*C* if the statement is false.

77. *Everyman* is classified as a morality play, because it deals with material drawn from the *Bible*.

78. The play is essentially an allegory concerning the way to salvation, because the characters are chiefly personifications of human qualities and the plot is an account of how *Everyman* makes his peace with God.

79. The author describes a systematized and predictable arrangement of the ways of God, because the steps necessary for the salvation of *Everyman* are clearly defined and commonly accepted.

Etc.

After each item number on the answer sheet, blacken space

*A* if the item at the left of the page is of greater magnitude than the item at the right.

*B* if the item at the right of the page is of greater magnitude than the item at the left.

*C* if the two items are of equal magnitude.

53. Amount of energy released in external respiration.	....	Amount of energy released in internal respiration.
--	------	--

54. Amount of time the ventricles of the heart are contracted.	....	Amount of time the auricles of the heart are contracted.
--	------	--

55. The rate at which blood pressure falls as the blood passes through the capillaries.	....	The rate at which blood pressure falls as the blood passes through the veins.
---	------	---

Etc.

ANSWER SHEET

After each item number on the answer sheet, blacken space

*A* if the event at the left occurred before the event at the right.

*B* if the event at the right occurred before the event at the left.

*C* if the events occurred at approximately the same time (within about a year of each other).

126. The beginning of our war with Mexico. .... The annexation of Texas.
127. The enactment of the Fugitive Slave Law. .... The Dred Scott Decision.
128. The attack on Fort Sumter. .... Fighting in Kansas between pro-slavery and anti-slavery groups.
- Etc.

There should be sufficient relationship between the paired events to warrant their being paired. Care should be taken in writing "C" items to give events that occurred simultaneously or very nearly simultaneously. Note the qualifying remark with respect to category "C" in the directions given above.

Exercises of the type illustrated below are useful in measuring how well students handle correlated or cause-and-effect relationships. In writing such items where the relationship is definitely cause and effect, the cause should be given first.

After each item number on the answer sheet, blacken space

*A* if increase in the first of the things referred to is accompanied by increase in the second, or if decrease in the first is accompanied by decrease in the second.

*B* if increase in the first of the things referred to is accompanied by decrease in the second.

*C* if the second of the things referred to remains constant, or approximately constant, even though the first increases or decreases.

188. Amount of carbonates dissolved in the water of a river.  
Number of clams in the river.
189. Temperature of the environment of a mammal.  
Body temperature of the mammal.
190. Number of lemming in an arctic area.  
Number of caribou in the same area.
- Etc.

One means of presenting students with exercises which require reflective thinking rather than memory alone is to set up hypothetical situations which differ from the situations encountered during instruction. The exercises given below illustrate such a series. The facts required in answering these exercises correctly were a part of the regular instruction in physical science.

You have acquired some knowledge of the earth and its motions as they really exist. In this exercise you are to identify the effects of some wholly imaginary conditions. After each item number on the answer sheet, blacken space

- A if the item would be true if the earth were *not* inclined on its axis.
- B if the item would be true if the orbit of the earth was a circle rather than an ellipse.
- C if the item would be true if the earth revolved toward the west rather than toward the east.
- D if the item would be true if the earth had half its present diameter but retained its present mass.
- E if the item would be true if the earth had no moon.

(Assume only one of the above imaginary conditions occurs at a time.)

- 44. All the solar days would be of equal length.
  - 45. Objects would weigh four times as much as they do now.
  - 46. The celestial equator and the ecliptic would be identical.
  - 47. The sun would set in the east.
- Etc.

The following series of exercises also present a novel situation. It is also possible that the uniqueness of such a series of exercises creates better examination rapport.

Suppose that Darwin, Weismann, Aristotle, DeVries, and Lamarck are broadcasting a round-table discussion from station STYX in Hades. Identify the speaker of each *numbered* statement by using the following code:

- A—Darwin
- B—Weismann
- C—Aristotle
- D—DeVries
- E—Lamarck

One of the group begins the discussion by saying, "A few days ago I read the following paragraph in a recently published textbook of biology:

"A population of slow-moving carnivores is established in a region in which plenty of food is available in the form of slow-moving herbivores. Unusual temporary conditions enable some of these carnivores to pass over a barrier into a new region. (Neither these animals nor their descendants ever pass back over the barrier to the old region.) In the new region the only possible food for the carnivores consists of fast-moving herbivores. Many thousands of years later the only descendants of the carnivores that

are living in the new region have characteristics that make them fast-moving.

"What is the explanation of this phenomenon?"

203. "Well, obviously the explanation is a simple one 'Chasing the fast-moving herbivores exercised the legs of the carnivores so that they became more effective means of locomotion. The offspring of these carnivores inherited better legs.'"
  204. "Impossible. As I was the first biologist to proclaim, germ cells come only from pre-existing germ cells—your remark implies that they can be influenced by the body cells."
  205. "I do not agree with the last remark. I had supposed that the germ cells are composites of substances (gemmules) derived from all organs and tissues of the body."
  206. "This discussion is too much for me, even though I credit myself with starting the idea of evolution."
  207. "As a matter of fact the correct explanation is that the variation which produced the fast-moving carnivores is due to mutations, or chance changes in the genes."
- Etc.

*Exercises pertaining to quoted material.*—The following directions are useful in writing classification exercises which pertain to quoted material (a paragraph or two, or even a graph).

Making your judgments only in terms of the information given above, classify each of the following items by blackening space

*A* if the item is *definitely true*.

*B* if the item is *probably true*.

*C* if the item is *definitely false*.

*D* if the item is *probably false*.

*E* if the information given is not sufficient to indicate any degree of truth or falsity in the item.

Probably-true statements are justifiable interpolations, extrapolations, or predictions from the information or data given. They may represent legitimate generalizations from information describing a sample or deductions with respect to a sample where the information pertains to things in general. For example, if the selection pertains to industrial conditions characteristic of the war years, a statement to be marked "B" may describe a trend or condition in a single industry analogous to

the general trend or condition, although the particular industry is not mentioned in the selection. Similarly, a statement to be marked "D" may describe a trend for a particular industry opposite to that of the general trend described in the selection. Such exercises may be scored not only to determine the general accuracy of the students in interpreting data, but also to identify students who have tendencies to be overcautious, to go beyond the data, or to make crude errors of judgment. For example, the over-cautious student marks A items B or E, and C items D or E. Conversely, the student who tends to go beyond the data, or over-generalize, marks B items A, D items C, and E items by some letter other than E.

Exercises of this type are very appropriate in achievement tests. Since the facts needed in thinking are presented in the quoted material, it seems legitimate to attribute differences in the scores of the students to differences in their ability to think in the subject-matter field. Where the quoted material is novel, or is on a level above that encountered in the course, the scores may be useful as predictions of future success in advanced courses in the same field. Since course marks should be predictions of future success rather than rewards for time spent in class, achievement tests used in determining final marks may well contain a considerable proportion of such exercises.<sup>2</sup>

*Exercises Involving Diagrams or Pictures.*—If there are no more than five pictures or things to be labeled in a diagram, the pictures or parts may be labeled A, B, C, D, and E, and the directions preceding the items may be written: "After each item number on the answer sheet, blacken the *one* lettered space which designates the part of the diagram (or the picture) to which the item correctly refers." Since the categories are inherent in the diagram or pictures no further directions are needed. The directions, the diagram or pictures, and the statements which constitute the test items should all be on the same page of the test booklet or on facing pages.

<sup>2</sup> For further information with respect to exercises of this general type, see: Smith, Eugene R., Tyler, Ralph W., and the Evaluation Staff. *Appraising and Recording Student Progress*. New York: Harper and Bros., 1942, 550 p.



The following directions are useful where comparisons are to be made between two diagrams or pictures, for example, reproductions of paintings representing different styles of art. In this instance the pictures may be labeled with Roman numerals.

After each item number on the answer sheet, blacken space

*A* if the item is true of picture I.

*B* if the item is true of picture II.

*C* if the item is true of both pictures.

*D* if the item is true of neither picture.

Where more than five parts of a diagram or locations on a map need to be identified, it is usually preferable to write exercises of the multiple-answer rather than the classification type. Each exercise may begin with a reference to one of the diagram symbols, for example, "The symbol 'I' refers to (A. . . . ., B. . . . ., C. . . . ., etc.)." In this case the order of the exercises should be the order of the symbols in the diagram. In some cases one can use the answer sheet or exercise numbers as diagram or map symbols. For example, suppose that a map exercise involves locations of battlefields of the Civil War. Number 126 on the map may be in the location of Gettysburg. The corresponding exercise, given on the same page, or on a facing page, and in a series of similar exercises may simply be:

126. A. Antietam, B. Gettysburg, C. Chancellorsville, D. Bull Run, E. Petersburg.

The directions for such a series of exercises may be phrased: "After the number on the answer sheet which corresponds to each map or exercise number, blacken the *one* lettered space which designates the correct answer."

Answer sheet numbers may also be used as symbols in a diagram when five or more than five parts need to be labeled, but there are no more than five possible labels or categories. In this case the following directions are useful:

After the number on the answer sheet corresponding to each number in the diagram, blacken space

*A* if the number refers to .....

*B* if the number refers to .....

*C* if the number refers to .....

*D* if the number refers to .....

*E* if the number refers to .....

Such directions are applicable to genetics exercises in biology where squares and circles representing male and female offspring are numbered and are to be classified according to genotype formula, or for example, as "a normal male," "a normal female," "a color-blind male," or "a female carrier." The square and circle representing the first mating are not numbered and are labeled in words. There may be more than one square or circle representing each of the types of offspring previously mentioned. The use of these directions need not be restricted to genetics exercises.

In certain fields it is effective to use small diagrams labeled A, B, C, D, and E as answers to ordinary multiple-answer exercises. For example: "Which of the following curves best represents the distribution of intelligence test scores of a large group of 12-year-old children?" "Which of the following symbols represents a cold front on a weather map?" "Which of the following tools should be used to . . . . .?" An effective E answer is "None of the above . . . ." which, if used in several of such exercises, should at least once be the correct answer.

After exercises have been written they should be carefully checked for accuracy of phraseology and correctness of the key. It is very desirable to have other teachers of the subject to evaluate the exercises with respect to fairness, freedom from ambiguities, and elimination of the too obviously incorrect answers. It is an excellent practice in seeking helpful criticism to have other teachers attempt to key the exercises. When teachers respond differently to an exercise, faults may be encountered of which the writer of the exercise was unaware.

When a test is ready for final typing some simplification of directions may be considered desirable. The directions for the first series of exercises may begin "After the number on the answer sheet which corresponds to that of each of the following items, blacken space . . . ." while the directions for the later series may commence "After the item number on the answer sheet . . . . .," or even more simply "Blacken space . . . . ."

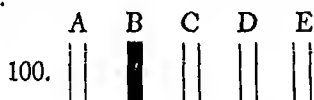
In the case of classification exercises, some simplification of categories may be desirable after teachers and students have

become familiar with exercises of this type. The categories may become a simple code as illustrated by the dialogue exercise on pages 368-369.

When a test is typed for planographing, or stencils are cut for mimeographing, care should be taken that the students will not need to turn back to a preceding page for directions or to refer to a diagram. This is particularly important in the case of classification exercises. It is less important in the case of exercises of other types.

General directions should be given at the start of each test with respect to careful marking of answer sheets. In the Chicago City Junior College the following directions are given at the start of each examination:

When marking your answers on the answer sheet, you must use a pencil filled with the special scoring machine lead. Make your marks thus:



1. Solid black marks are made by going over each mark two or three times and by pressing firmly on your pencil.
2. If you change your mind, erase your first mark completely.
3. Make no unnecessary marks in or around the dotted lines. Do not rest your pencil on a lettered space while deciding which space to mark.
4. Keep your answer sheet on a hard surface while marking your answers
5. Make your marks as long as the pair of dotted lines.
6. Make *one* mark and only *one* mark after each answer sheet number. Your score will be reduced by an amount equal to the number of extra marks on your answer sheet.

Each examination concludes with the following statement:

When you have finished the test go over each of your marks again with the special pencil to make each mark solid, black, and glossy.

**YOUR SCORE WILL BE REDUCED BY AN AMOUNT  
EQUAL TO THE NUMBER OF EXTRA MARKS ON  
YOUR ANSWER SHEET.**

When proctoring a test which is to be machine scored, inspect each student's answer sheet soon after the students begin

work. Insist on glossy, black marks and the complete erasure of superfluous marks. Make a second inspection shortly before the students finish. There should be one mark only after each answer sheet number. Be sure each student uses a pencil containing the special lead. Accurate and rapid machine scoring is impossible unless these precautions are observed.

The preceding discussion should prove helpful in training teachers to become expert in the art of writing exercises for machine-scored achievement tests. Critical evaluation of exercises by teachers is suggested above. An examination staff can promote such evaluation by subjecting tests which have been given to item analysis. A simple type of such analysis has been described by the author<sup>3</sup> and a more complex but more adequate type by Davis.<sup>4</sup>

The item analysis data pertaining to individual exercises may be recorded on the original cards containing the test items, or they may be recorded adjacent to each item in a keyed test booklet. The per cents of correct response are particularly useful to teachers in evaluating attainment of objectives. Both the per cents of correct response and the item test correlations are useful in identifying faulty exercises. A combination of scholarly writing of exercises, painstaking criticism and editing prior to first use, and thoughtful interpretation of item analysis data cannot help but result in the production of superior achievement tests.

---

<sup>3</sup> Engelhart, Max D., "How Teachers Can Improve Their Tests," *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, IV (1944), 109-124.

<sup>4</sup> Davis, Frederick B., "Item Analysis Data, Their Computation, Interpretation, and Use in Test Construction," *Harvard Education Papers* Number 2, Cambridge, Massachusetts: Harvard Graduate School of Education, 1946. 42p.

# VOCATIONAL INTERESTS AND VOCATIONAL CHOICE: PRESENT KNOWLEDGE AND FUTURE RESEARCH IN THEIR RELATIONSHIPS

DONALD E. SUPER

Teachers College, Columbia University

IN 1931 Fryer reviewed the published studies of vocational interests and vocational choices, with rather discouraging implications for their study, prediction, and control (7). In 1943 Strong published his report of twenty years of work in the measurement and analysis of vocational interests (21), a report which the writer has elsewhere (23) characterized as a milestone in the field of vocational psychology. In a current review of work with the *Kuder Preference Record*, the writer has in still another place (24) pointed out the great progress which has been made in the study of interests. It therefore seems timely to survey what we now know about the relationship of vocational interests and vocational choice, in order to ascertain the nature of needed research.

## *Expressions and Manifestations of Preferences and Interests*

In considering the relationship of vocational interests to vocational choices, both terms need to be defined: there has been considerable confusion in the literature because of failure to make adequate distinctions, illustrated in Lehman and Witty's early criticisms of Strong's Blank (12) and in Fryer's pioneer volume (7). An occupational choice may be *expressed*, that is, stated in response to a question: As numerous investigators have shown that the phrasing of the question has an important bearing on the amount of realism in the response [Gilger (8), Lurie (13), Trow (26)] one should differentiate

between expressed *choice*, *preference*, and *fantasy*. Or a choice may be *manifested*, that is, demonstrated by participation in an occupation or in preparation for it. It might seem that participation would be an objective and valid index of interest or preference, but Stunt (22) and Vernon (28), for example, have shown that many students in teachers' colleges are there for reasons other than interest in teaching as an occupation and even without any intention of entering that occupation, and various other occupations are not infrequently used as means of entry into still other fields (17). A vocational interest, similarly, may be an *expressed* interest, one indicated by a statement in response to a question. As in the case of choices, the phrasing of the question affects the answer. It may be *manifested*, as in the case of choice, through participation—in which case the two are synonymous—or *tested* by some objective performance such as demonstrated knowledge of the vocabulary, tools, or other appurtenances of the occupation, in which case manifest choice and interest might not be identical. Or, finally, interest may be *inventoried* by means of an interest blank such as Strong's, in which the interests of men engaged in various occupations are used as the norm, or by means of blanks such as Kuder's and Allport and Vernon's in which the relative strength of liking for various types of activities is measured.

Strong's work has focused on inventoried interests, although he has related these to expressed choice and to choice as manifested in the occupation entered. In this direct way, then, his work bears on the problem of vocational choice. It also bears on it indirectly, because of the role interest plays in choice: presumably anything found out about interests through Strong's approach has at least some bearing on choices, which they help to determine. This will in time be true of Kuder's work also, after more studies using his inventory are completed.

#### *Correlation between Inventoried Interest, Expressed Choice, and Manifest Choice*

Five studies using Strong's Men's Blank have reported significant correlations between expressed or manifest choice of occupation and interest as inventoried by Strong's Blank, four

using the women's form have found negligible relationships, and one with men has found that the extent of the relationship depends upon intelligence. Carter, Taylor, *et al.* (4, 5, 25) studied the interests and preferences of tenth-grade students, finding that 65 per cent of the group had letter ratings higher than C in the field of their choice on Strong's Blank. Dyer (6) followed up a group of 101 college students ten years after graduation, and reported that 66 per cent of the group who were employed in the college-preferred vocation ten years later rated A or B+ in the appropriate occupation on Strong's Blank. Strong (21, 393 ff.) followed up two different groups of Stanford students nine and ten years after graduation. His two studies are in substantial agreement, although in the first study the testing was done with freshmen, and in the second it was done with seniors, the agreement between choice and inventory score being somewhat closer for the more mature group. Of the 197 men tested as seniors, 116 had not changed the occupational preference expressed in college when compared with a differently defined type of "choice" (actual occupation of graduates) ten years later; occupational data for 95 were comparable to Strong's scales, and for these there was no appreciable change in interest scores over the ten-year period. Those who were not sure of their choices in college were from 7 to 11 standard score points lower than those who were sure. Of the 81 whose expressed choices (occupational preferences) in the senior year in college and manifest choice (actual occupation) ten years later were not identical, 63 had choices which could be compared to scores on Strong's scales. The scores of this group had changed, on retesting, ten years after graduation, the mean for abandoned occupational preference being significantly lower than it had been when it was the preferred vocation in college. Similarly, the score for the actual occupation to which these men had changed during the ten years after graduation was significantly higher than it had been as an unchosen occupation in college. Van Dusen (27) has reported similar results. Strong explains this finding on the basis of increased understanding and allegiance to a self-concept. Perhaps, also, responding to a stereotype plays a part, as observed by Paterson (15). Moffie

(14) studied college students, obtaining self-ratings of interest in Strong's occupational groups from some 80 subjects and correlating these with Strong's scales; the coefficients ranged from .05 to .54. Apparently these students had better insight into their interest in some fields than in others, which may be due to the effect of social distance and lack of information in a socially selected group (NYA students) rather than of inadequate choices, for the experiment did not involve expressions of first preference but rather rankings of certain occupational groups.

Lalager (11) tested 703 high-school juniors with Strong's Women's Blank, a less discriminating instrument than the men's form, and found little agreement between Strong's scales and expressed occupational choice. Skodak and Crissy (19) also found the women's scales not closely related to expressed choices and less "useful" than the latter because of the concentration of scores in certain occupations involving an interest factor common to many women. Stuit (22) found that only about 10 per cent of the women freshmen in a teachers' college ("teachers" by manifest choice) rated A or B+ on the Strong teachers' scales. And Bedell (1), studying a similar group, reported correlations of more than .50 between only two of seventeen women's scales and occupational preferences.

Wrenn (29, 30) studied the intelligence and vocational preferences of nearly 10,000 California junior-college students, then administered Strong's Blank to the top 5 per cent and bottom 15 per cent of the intelligence distribution. Of the high group, 45 per cent rated A on the scale for the occupation of their choice, as contrasted with 22 per cent of the "low" intelligence group. This is what one would expect a priori: the more able students understand their interests and so make better choices, and probably also have more adequate occupational stereotypes to guide them in responding to the items in Strong's Blank.

*Relationship between Change of Interest, Maturation,  
and Experience*

The effects of maturation on interests have been studied cross-sectionally by Strong (21: Ch. 13) and longitudinally by



Carter and associates (4, 5, 25) in connection with the California Adolescent Growth Study. Strong's data were obtained from college students and adults, all males; his method consisted of tabulating the specific likes and dislikes of men aged 15, 25, and 55, and of comparing the percentages in each group indicating liking, indifference, or dislike for each occupation, school subject, activity, etc., in his Interest Blank. The inventories of these groups were also scored with the standard occupational keys. These analyses showed that older adolescents tend to be more interested in scientific and technical matters, whereas the interests of young men tend to be more social, that is, they tend to stress human values more. Older men have interests very much like those of younger men, though less concerned with athletics and somewhat more restricted in nature. These findings might be the result of cultural differences in generations, but Strong interprets them as age differences, and this is confirmed by Carter's work, which found the same interest trends in adolescents studied genetically during their four years in high school, these interest patterns beginning to crystallize at about age 14. In other words, the differences found by Strong between older adolescents and adults were measured by Carter as trends in adolescents as they matured. Strong attributes these changes to endocrine development and sexual maturation.

The effect of school experiences on inventoried interests has been investigated by Burnham (3), Klugman (10), and Van Dusen (27). Klugman retested a group of 207 high-school girls after one year of study in the commercial course, and found no significant relationships between their change of clerical interests and their grades in commercial courses. This was a poor choice of group for such a study, however, in view of the general tendency for girls and women to make high scores on the clerical scales. Burnham's study is of more interest, as he worked with men college students and found little relationship between change of interest and college grades. Van Dusen worked with engineering students at the University of Florida but used claimed interest rather than grades as his criterion. He found that the inventoried interests of students whose claimed vocational interest changed while in college did not change, whereas

the inventoried engineering interest of those whose expressed choice remained in engineering did increase significantly. Perhaps the former were an immature group, who discovered as they studied engineering that theirs was a general interest in science, and the latter an emotionally more mature group whose interest in technical matters deepened and broadened with their exposure to the field. Whatever the explanation of Van Dusen's findings, Strong did not confirm them in a study of engineering students at Stanford (21, pp. 277 ff.): he found no increase in the inventoried engineering interests of men who continued in that field until graduation. Various possible explanations suggest themselves, ranging from the degree of occupational orientation characterizing entering freshmen at the two universities in question (one near industrialized areas and the other remote) to differences in selection procedures in the two engineering schools in question, but data by which to judge these factors are not available.

Strong followed the students referred to above, together with others in various curricula, into their subsequent careers, retesting them five years after graduation. While significant relationships were found between their interests and their stability in or change of occupation, there was little change in their interest scores. He was not focusing on this latter problem, but the data available and the incidental analysis indicate that inventoried interests are rather stable and that at least in socially mobile adults it is choice and experience which vary to agree with inventoried interests rather than inventoried interests which vary as a reflection of choice and experience. The implications of this conclusion are so important that further intensive studies of this problem should be made in order to verify it and to ascertain at what ages experience does affect inventoried interests to an appreciable degree, how it influences them, and the age at which these influences cease to have significant effects. Carter's studies, previously discussed, have thrown needed light on the role of maturation in the basic or inventoried interests of adolescents; comparable work remains to be done on the role of experiences of various types at that age.

There have been a number of experimental studies of the effect of specific experiences on expressed preferences (2, 9, 16, 18), but these have involved "experiences" which were so trifling as to be meaningless, were so poorly designed as to prove nothing, or were extremely limited in their scope. They throw no light on the role of experience in the development of interests.

### *Conclusions*

The findings reviewed in this paper appear to have the following implications for further investigations of the determinants of vocational choice and the development of interests:

1. Little is known of the role of experience in the development of vocational interests in the formative period: adolescence. This is the outstanding problem still to be studied.
2. Studies of this problem might be longitudinal or experimental, or both. In any case, the quality of experience needs to be carefully evaluated.
3. The definitions of the term "choice" must be clear and their implications for the design of experiments understood.
4. An adequate interest inventory provides a check on the intrinsic appeal of the chosen occupation to the person choosing it, and will thus help to throw light on the relative importance of extrinsic factors such as family, economic conditions, etc., but is somewhat affected by stereotyping.
5. The subjects should be males, in order that a meaningful measure of interest may be obtained.
6. The subjects should include a relatively able group, capable of understanding the occupational significance of their interests and of manifesting these interests in appropriate occupations.
7. At the same time, less able groups may be needed in the same study in order that relationships between choice and ability, interest and ability, etc., may not be hidden by a restriction of range.
8. The experiences of the subjects prior to the initiation of the study need to be analyzed, in order to ascertain what influences have been at work on them so far, e.g., the breadth and nature of previous occupational experiences, whether first-hand or vicarious, affects preferences and perhaps even basic interests.

9. Selective factors affecting the sample studied must be known, e.g., if students in one college are studied, what is found true for them may not apply to students in another college if the maturity of the student is one of the criteria considered by the admissions office of one college but not by that of another.

10. A vocational interest inventory such as Strong's or Kuder's should be used in studying choices as well as preferences, because inventoried interests appear to be more stable than vocational choice in adolescents and young adults, whether choice is defined as preference or as the acceptance of employment. This seems to be due to the importance of maturation, and to the relative unimportance of experience in adults and possibly in college students; as stated above, the role of experience in adolescence is virtually unknown.

#### REFERENCES

1. Bedell, R. "The Relationship Between Self-Estimated and Measured Vocational Interest." *Journal of Applied Psychology*, XXV (1941), 54-66.
2. Bunting, J. R. "Counseling Alters Pupils' Choice." *Occupations*, XVIII (1939), 174-176.
3. Burnham, P. S. "Stability of Interests." *School and Society*, LV (1942), 332-335.
4. Carter, H. D. and Jones, M. C. "Vocational Attitude Patterns in High-School Students." *Journal of Educational Psychology*, XXIX (1938), 321-334.
5. Carter, H. D. and Taylor, K. v. F. "Strong Vocational Choices and Interest Test Scores on High-School Students." *Journal of Psychology*, XI (1941), 297-306.
6. Dyer, D. T. "Relation Between Vocational Interests of Men in College and Their Subsequent Occupational Histories for Ten Years." *Journal of Applied Psychology*, XXIII (1939), 280-288.
7. Fryer, D. *Measurement of Interests in Relation to Human Adjustment*. New York: Henry Holt and Company, 1931.
8. Gilger, C. A., Jr. "Declaration of Vocational Interest." *Occupations*, XX (1942), 276-279.
9. Kefauver, G. N. and Hand, H. C. *Appraising Guidance in Secondary Schools*. New York: Macmillan Company, 1941.
10. Klugman, S. F. "Permanence of Clerical Interest in Relation to Age and Various Abilities." *Journal of Social Psychology*, XXI (1945), 115-120.
11. Laleger, G. E. "Vocational Interests of High-School Girls." *Teachers' College Contributions to Education*, No. 857, 1942.

12. Lehman, H. C. and Witty, P. A. "Vocational Counseling—The Interest Inventory." *American Journal of Psychology*, XLIV (1932), 801-805.
13. Lurie, W. A. "Estimating the Level of Vocational Aspiration." *Journal of Social Psychology*, X (1939), 467-473.
14. Moffie, D. J. "Validity of Self-Estimated Interests." *Journal of Applied Psychology*, XXVI (1942), 606-613.
15. Paterson, D. G. "Vocational Interest Inventories in Selection." *Occupations*, XXV (1946), 152-153.
16. Recktenwald, L. N. "Attitudes Toward Occupations Before and After Occupational Information." *Occupations*, XXIV (1946), 220-223.
17. Reinhardt, E. "Vocational Expectations of Freshmen in a Teachers' College." *School and Society*, XLIV (1936), 518-520.
18. Remmers, H. H. and Whisler, L. D. "Effects of a Guidance Program on Vocational Attitudes." *Purdue University Studies in Higher Education*, III (1938), 68-82.
19. Skodak, M. and Crissey, O. L. "Stated Vocational Aims and Strong Interest Scores of High-School Senior Girls." *Journal of Applied Psychology*, XXVI (1942), 64-74.
20. Strong, E. K., Jr. "Predictive Value of the Vocational Interest Test." *Journal of Educational Psychology*, XXVI (1935), 331-349.
21. Strong, E. K., Jr. *Vocational Interests of Men and Women*. Stanford University: Stanford University Press, 1943.
22. Stuit, D. B. "A Study of the Vocational Interests of a Group of Teachers and College Freshmen." *Journal of Applied Psychology*, XXII (1938), 527-533.
23. Super, D. E. "Strong's Vocational Interests of Men and Women" (a review). *Psychological Bulletin*, XLII (1945), 359-370.
24. Super, D. E. "The Kuder Preference Record in Vocational Diagnosis and Guidance." *Journal of Consulting Psychology* (in press).
25. Taylor, K. v. F. "Reliability and Permanence of Vocational Interests of Adolescents." *Journal of Experimental Education*, XI (1942), 81-87.
26. Trow, W. C. "Phantasy and Vocational Choice." *Occupations*, XX (1941), 89-93.
27. VanDusen, A. C. "Permanence of Vocational Interests." *Journal of Educational Psychology*, XXXI (1940), 401-424.
28. Vernon, P. E. "Educational Abilities of Training College Students." *Journal of Educational Psychology*, XI (1939), 233-250.
29. Wrenn, C. G. "Aiding the Fit." *Journal of Higher Education*, VI (1935), 357-363.
30. Wrenn, C. G. "Intelligence and the Vocational Choices of College Students." *Educational Record*, XVI (1935), 217-219.



## THE SELECTION OF FOREMEN<sup>1</sup>

MILTON M. MANDELL

THIS article is a report of an examination program for the selection and promotion of five levels of foremen in shipyards, air stations, ordnance plants, warehouses, and other field installations of the Navy Department. While these units are government owned and operated, they furnish basically the same type of selection problems for supervisory positions as are found in similar industrial activities under private ownership. One can reasonably extend this statement, because of the variety of occupations found in these Navy establishments, and say that the selection principles found to be worth while in these activities would probably also be appropriate for most supervisory positions in American industry.

The selection of supervisors in American industry and government is of paramount importance. Practically the entire impact of management policies falls on first- and second-level supervisors. The day-by-day interpretation and administration of these policies is in their hands. Ineffective supervision makes meaningless the best policies, the best technological equipment, and the efforts of the best workers. Yet one can review the whole literature of testing and find but a small group of studies on the development and validation of tests for the selection of these key workers. The development of clerical, mechanical, and sales aptitude tests is often described, but most of the literature on supervisory, as on administrative, selection represents merely an attempt to identify the factors that are important in supervisory selection. One may venture the opinion that adequate attention to supervisory selection would immeasurably increase the effectiveness of present methods.

---

<sup>1</sup> Appreciation is due to Thomas L. Bransford for his sponsorship and supervision of this program, to Marjorie West, Sidney Adams, Jean Hussey, and Ernest Primoff for their test construction work, and to Meyer Shultz, Jeanne Davis, and B. J. Winer for their statistical work.

More than 10,000 supervisory positions are included in this selection and promotion program. The program requires that only those who are among the top employees on a ranked list may be appointed to any supervisory position. The examination on which this ranking is based includes, for the first and second levels (the two lowest supervisory grades), a written test, a numerical evaluation of the training and experience of the employee, and his efficiency rating. These three ratings are combined into one for the purpose of obtaining a final rating. For the three higher grades of supervision, an oral interview, which provides a numerical rating, is an additional part of the examination.<sup>2</sup> The written test has a weight of 50 per cent for all five levels, and the remaining 50 per cent is divided equally among the other parts of the examination. There are two parts to the written test: Part I is used for all five levels, while Part II is used for the three higher grades only. Part I is in 5-choice form while Part II contains questions of the completion and short-essay types. Part I, which is sometimes referred to as the Supervisory Judgment Test, emphasizes questions on the foreman-worker relationship and questions on specific supervisory techniques such as placement, training, safety, and performance evaluation. The questions on the Supervisory Judgment Test are general in that they apply to all supervisory jobs included in this program and are not specific as to any occupation or industry. Great care has been taken to make these questions as realistic and as unacademic as possible. Part II of the written test, which is limited to the three higher supervisory grades because the comparatively large number of examinations in the lower grades do not permit this additional examining work, is tailor-made for the particular occupation or shop for which the examination is being given. This part attempts to measure breadth of knowledge and judgment in relationship to the super-

<sup>2</sup> The supervisory titles in use in Navy field establishments for these positions are, starting from the bottom: Leadingman, Quartermen, Chief Quartermen, Foreman, and Master. These levels will be referred to as first-, second-, third-, fourth-, and fifth-level supervisors, respectively. The Leadingman is a full-time, rather than a "working," supervisor who has journeymen, helpers, and laborers in his group. The total number of employees under his supervision would be roughly 20 or 25. The Master, at the other extreme, is the head of a major shop. At the present time, Masters supervise as many as 1,000 employees; at the peak of the war, some Masters supervised as many as 10,000 employees.



visory and administrative duties of the particular position for which the examination is being given. The following directions are used to guide Navy and Military specialists who draft the material for Part II of the written test:

Approximately fifteen typical tasks should be prepared for each trade or shop for which an examination is needed. For example, the task might be. "Manufacture 250 3" valves, composition, 200#, steam working pressure, flanged, Navy Standard." The following precautions should be observed in preparing these tasks:

- (1) The tasks should be representative of all functions of the trade or shop. For example, in a naval shipyard the significant categories might be (a) new construction and repair; (b) shop and field work; and (c) types of vessels . . . All significant phases of activity of the trade or shop should be covered in the questions that are prepared.
- (2) The tasks should be sufficiently complete so that the applicant does not have to guess as to what is intended. A contractor could not make a cost estimate unless he had detailed information as to the type and size of house desired, types and quality of material desired, etc. In the above example, the number, size, and type of valves to be manufactured had to be furnished in order to give the candidate sufficient information to answer the questions. Other types of information that may be necessary are: (a) the amount of time available to do the job; (b) the status of the job before the work is started on it by the shop or trade; (c) the type of ship or aircraft on which the work is to be done; etc. . . .
- (3) The tasks should be such that the following questions can be asked in regard to each of them. If any of the following questions are not appropriate to any task submitted, please indicate this fact when transmitting the material. Indicate which question or questions are not related to the task, and, furthermore, check the information you have included in the description of the task against this list of questions to make certain that enough information has been given to insure that the task is not ambiguous.
  - (a) How many man-days, by trade and rating, would be required to perform this work?
  - (b) List the types and quantity of equipment that would be needed to perform this work.
  - (c) List the safety hazards that might occur in performing this work.

- (d) What other shops are involved in this work? Indicate the shops that do their part of the work before, during, and after the shop for which the examination is being given.
- (e) List the major technical changes that have been developed in performing this work during the past five years and give the major advantages of each change.

The use of competitive examinations for the selection of supervisors in Navy installations has been in operation since about 1920. It has the support of the workers and supervisors because it provides an objective method for selection. It has the support of top-management for the same reason. To the extent that this program tends to eliminate charges of favoritism, it is a useful morale builder. And to the extent that it furnishes a more valid method for the selection of supervisors than the method more commonly used, it improves administration at a key point. The remainder of this article will describe a study of the validity of various types of written tests that might be included in this examination program.

Any description of the results of a validation study should include, to be meaningful, the institutional situation in which it was conducted. The importance of the validity coefficients obtained in such a study should be interpreted in the light of such information. The following subjective conditions affected the results of this study, some positively, some negatively. In the first place, the trial groups attempted to do their best work in answering the questions. They were told that this was a test of the tests and that the only purpose of the study was to determine which tests should be used in the future for actual selection. Code numbers rather than names were used on the answer and rating sheets. Also, the support of top-management had initially been obtained by extensive discussion of the purposes and methods of the study. The above condition helped to increase the value of the study.

The second important factor which should be mentioned affected the validity coefficients by lowering them, one can presume, although the extent to which the coefficients were lowered could not be measured easily since, because of the nature of the

criterion used, its reliability could not be estimated satisfactorily. Very extensive layoffs had occurred and more were expected in the near future in many of the activities where these supervisors worked. These layoffs had made all of the men jittery. It can therefore be presumed, although data are lacking, that the quality of the ratings obtained was lowered as a result of this uncertainty. While the men had been assured that neither the ratings nor the test scores would affect their status in any way, their positions were in jeopardy because of the end of the war. This fact undoubtedly affected their willingness to rate their subordinates and colleagues objectively. This condition was known at the time the study was planned. It was decided to go ahead with the study because there was need for an immediate determination as to which types of tests were most valid. It was assumed, perhaps incorrectly, that only the absolute values of the validity coefficients would be affected by this condition, and not their relative validity.

### *The Population*

The groups included in the study were supervisors at all five levels in two shipyards, two supply depots (warehouses), and two air stations. In addition to the five levels of supervisors, a sixth level, that of working supervisor, was also included. These "straw bosses" do not have full supervisory status, and, unlike the full-time supervisors who are paid an annual wage, are paid on an hourly basis (10¢ more than the highest rate of the men they supervise). All supervisors were invited to participate in the study but it was emphasized that this was completely voluntary. A comparison of two large groups of participants and non-participants (absent either voluntarily, or because of sickness, or because of conflicting work schedules) indicated that they were approximately equal in age distribution. The number of supervisors at the two higher grades who did not participate was relatively larger than the number of the lower grades who did not participate. For this reason and also because of the limited number of supervisors at the two highest grades, the data presented are, in general, for the first three supervisory levels only.

The supervisors came from the three types of Navy activities described above and from a variety of trades. Included were supervisors of laborers, stevedores, plumbers, aircraft mechanics, sheet-metal workers, carpenters, power-plant operators, electricians, machinists, welders, automobile mechanics, boiler-makers, molders, riggers, pipefitters, patternmakers, copper-smiths, and supervisors from a host of other trades. The occupations range, as will be noted, from the lowest levels to the highest levels of skill. The study was deliberately designed in this manner because the objective of the study was to determine which test or combination of tests would measure the common supervisory elements of all of these positions, leaving to later determination the question as to which tests would best measure the special factors in each of the types of positions.

The size of the groups on which the data are based vary from one grade to the next and between Navy installations. The over-all conclusions at the end of this report are based on about 1,000 cases.

### *The Tests*

The statement has previously been made that the attempt in this study was to determine which tests would work effectively at all grade levels and for all occupations.<sup>3</sup> To achieve this objective, the following tests were included in the study:

1. *Supervisory Judgment*.—This 50-item, 5-choice, multiple-choice test attempts to measure two parts of the supervisor's job: (1) judgment on the problems of employee-supervisor relationships, and (2) judgment on problems of supervisory techniques such as performance evaluation, training, safety, and employee utilization. The test is, to the greatest extent possible, non-factual in that an attempt has been made in preparing the questions to make a high score impossible unless, in addition to the facts, the candidate has the judgment necessary to apply the facts properly to a given problem and to come out with the right answer. There is also a second part to the definition of "non-factual." In this sense, "non-factual" means the same as

<sup>3</sup> The same passing points will not be used for all supervisory grades and occupations. The data obtained in the study justify, and the methods of administration provide, successively higher passing points for the three lowest grades of supervisors.

"non-academic," that is, the questions are of such character that formal training in human relations or supervision is *not* essential to satisfactory performance on a test.<sup>4</sup>

In order to get the support of the men involved for the testing program it was considered essential that the questions should pose real problems, and not be merely academic exercises. While it is obviously impossible to meet this goal fully, the reaction of the men who participated in the study indicates that this goal was substantially achieved.

2. *Mechanical Principles*.—This test, consisting of 25 items, is similar to Bennett's *Test of Mechanical Comprehension*. This type of test had proved useful for supervisory selection previously although the reasons for this relationship are not definitely known.<sup>5</sup>

3. *Spatial Relations*.—This test consists of 25 paper form-board items. This type of test has been found useful for the selection of journeymen and it was included in this study to determine its usefulness for supervisory selection.

4. *Rules and Regulations*.—This test consists of 20 factual items on Navy Department and Civil Service Commission personnel regulations. Part of the supervisor's duty is to apply these regulations correctly and to answer the questions of his men. It was therefore thought appropriate to include items of this type.

5. *Reading Comprehension*.—This test consists of 20 items on the ability to read and understand quotations from general supervisory and technical literature. The large number of documents that a supervisor has to read made it seem desirable to include items of this type.

6. *Numerical Relationships*.—This test consists of 15 items containing numerical series, with the candidate supplying; as his answer, the number that should be next in the series. This test was included because it was assumed that the quantitative fac-

---

<sup>4</sup> No assumption has been made that training in supervision is not desirable; rather, the assumption that has been made is that one may learn to be a good supervisor by observation, experience, and by using intelligence, as well as by formal training.

<sup>5</sup> A summary of the results obtained in the use of the Bennett and other tests for supervisory selection is to be found in Mandell, Milton, "Testing for Administrative and Supervisory Positions," *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, V (1945), 217-228.

tors measured by the items would have some relationship to the calculations necessary in a supervisor's job.

7. *Blueprint Reading.* This test consists of 25 items that attempt to measure knowledge of blueprint terms and judgment in interpreting blueprint directions. In order to simplify the process of test preparation and administration, only ship blueprints were used. If possible, it would have been desirable to prepare questions on blueprint reading appropriate to building maintenance work and airplane mechanics.<sup>6</sup>

### *The Criterion*

The ratings were furnished by the supervisors and colleagues of those men participating in the study. The two types of rating forms used were graphic and rank-order. The instructions to the raters provided that they would rate those supervisors whose work performance they knew and whom they were willing to rate accurately. Both parts of this definition were stressed. Each man rated about six participants.

Inspection of the rating data received indicated that, in many cases, there was substantial disagreement among the raters and disagreement between the results obtained from the two rating methods. Cases were eliminated when these disagreements were substantial. There were conflicting objectives in making these eliminations. On the one hand, to get the highest reliability, it would have been desirable to eliminate all cases with more than a minimum of disagreement. On the other hand, in order to meet the above objective, the number of cases in the study would have been reduced to a level below that required for significant results. Therefore, the method used to eliminate cases kept both objectives in mind, and resulted in the elimination of about 25 per cent of the cases. The elimination of these cases improved the agreement among the raters somewhat. In other validity studies, in order to obtain maximum agreement, as many as 75 per cent of the cases have been eliminated.

<sup>6</sup> It will be noted that many, if not all, of the tests have too few questions for a reliable test. The administrative decision was made that it would be desirable, for purposes of the study, to try-out as many item types as possible and then to lengthen, in actual use, those tests which give the greatest validity.

*Statistical Data*

All of the correlations are Pearson product-moment correlations between test scores and job performance. In all cases, the correlations are uncorrected for any cause. They are based on the original test content.

## 1. SUPERVISORY JUDGMENT TEST

<i>Level</i> <sup>7</sup>	<i>Activity</i> <sup>8</sup>	<i>N</i>	<i>r</i>	<i>Level</i>	<i>Activity</i>	<i>N</i>	<i>r</i>
CQ and Q	NAS—Alameda	28	.34	L	NAS—Alameda (AR)	24	.48
Q	NSD—Oakland	20	.50	L	NSD—Oakland	47	.05
CQ and Q	NS—San Francisco	57	.06	L	NS—San Francisco	142	.08
Q	NS—Mare Island	68	.12	L	NS—Mare Island	171	.18
Q and L	NSD—Norfolk	60	.40	L	NS—Mare Island (Portion)	71	.36
L	NSD—Norfolk	46	.32	S	NSD—Oakland	41	.08
L	NSD—Norfolk (Portion)	39	.37				

It will be noted that seven of these thirteen correlations are above .30 and three of them are .40 or above. The range of these correlations is great, from .50 to .05. While a number of explanations may be made of these differences, two are offered: (1) the differences in reliability of ratings obtained at the various activities, and (2) the differences in variety of trades in each group. In other words, the more homogeneous the group, the more uniform are the standards used for rating. This second point is strikingly illustrated in the case of the two correlations reported for Leadingman at the Mare Island Naval Shipyard. Using the total group of 171 Leadingmen, who represent at least 20 different trades, the validity coefficient is  $.18 \pm .07$ ; using only the 71 Leadingmen in four of the highest skilled trades, the validity coefficient is  $.36 \pm .10$ . In other words, it would have been desirable to analyze the data separately for the supervisors of each trade, but the lack of sufficient cases made necessary the lumping together of a large number of trades. In actual practice, the examinations are administered separately by trade.

<sup>7</sup> The following abbreviations have been used for supervisory levels: F—Foreman; CQ—Chief Quartermaster; Q—Quartermaster; L—Leadingman; and S—Snapper (straw boss). In a few cases two levels of supervision have been combined in order to increase the N.

<sup>8</sup> The following abbreviations are used. NAS—Naval Air Station, NS—Naval Shipyard; NSD—Naval Supply Depot. The following terms are used for parts of an activity. AR—the Assembly and Repair Department, Public Work-maintenance activities; and Portion—a part of a major activity which is relatively homogeneous in terms of occupations represented.

## 2 MECHANICAL PRINCIPLES TEST

Level	Activity	N	r	Level	Activity	N	r
F and Q	NAS--Norfolk	59	.11	L	NAS--Norfolk	153	.16
Q	NAS--Norfolk	57	.09	L	NAS--Norfolk (AR)	129	.18
CQ and Q	NAS--Alameda	28	.06	L	NAS--Alameda (AR)	24	.30
Q	NSD--Oakland	29	.15	L	NSD--Oakland	47	.00
CQ and Q	NS--San Francisco	57	.22	L	NS--San Francisco	142	.09
Q	NS--Mate Island	68	.25	L	NS--Mate Island	171	.13
Q and L	NSD--Norfolk	60	.11	L	NS--Mate Island (Porton)	71	.25
L	NSD--Norfolk	46	.07	L and S	NAS--Norfolk (Public Works)	23	-.06
L	NSD--Norfolk (Porton)	39	.10	S	NSD--Oakland	41	.39

The validity coefficients for this test are lower, in general, than those for the Supervisory Judgment Test, although for Quartermen and Snappers at the Oakland Naval Supply Depot the correlations are .53 and .40. Combining this test with the Supervisory Judgment Test adds very little to the validity.\*

## 3 SPATIAL RELATIONS TEST

Level	Activity	N	r	Level	Activity	N	cr
F and Q	NAS--Norfolk	59	.18	L	NAS--Norfolk	153	.24
Q	NAS--Norfolk	57	.15	L	NAS--Norfolk (AR)	129	.31
CQ and Q	NAS--Alameda	28	.01	L	NAS--Alameda (AR)	24	.16
Q	NSD--Oakland	29	.51	L	NSD--Oakland	47	.04
CQ and Q	NS--San Francisco	57	.14	L	NS--San Francisco	142	.18
Q	NS--Mate Island	68	.25	L	NS--Mate Island	171	.08
Q and L	NSD--Norfolk	60	.04	L	NS--Mate Island (Porton)	71	.20
L	NSD--Norfolk	46	-.10	L and S	NAS--Norfolk (Public Works)	23	.47
L	NSD--Norfolk (Porton)	39	.04	S	NSD--Oakland	41	.37

The zero-order correlations for this test are lower than for the Supervisory Judgment Test; however, this test had higher correlations than most of the other tests for Leadingmen as a whole. This adds some significance to these correlations because the population in the Leadingmen groups was much larger than for the higher supervisory levels.

The results with this test were generally poor for Leadingmen and relatively good for the higher grades. The relatively low reliability of the test, .459, indicates that if the reliability

\* In general, combining tests by the Wherry-Gaylord test selection method, limiting the individual test weights to positive integers, yielded correlations only slightly higher than the highest single test correlations. Combining tests by the usual multiple-regression method resulted in multiple correlations which were somewhat higher.



## 4 RULES AND REGULATIONS TEST

Level	Activity	N	r	Level	Activity	N	rr
F and Q	NAS—Norfolk	39	.30	L	NAS—Norfolk	153	.07
Q	NAS—Norfolk	37	.28	L	NAS—Norfolk (AR)	129	.09
CQ and Q	NAS—Alameda	28	.42	L	NAS—Alameda (AR)	24	.09
Q	NSD—Oakland	20	.62	L	NSD—Oakland	47	.08
CQ and Q	NS—San Francisco	57	.11	L	NS—San Francisco	142	.09
Q	NS—Mare Island	68	.15	L	NS—Mare Island	171	.20
Q and L	NSD—Norfolk	60	.19	L	NS—Mare Island (Portion)	71	.16
L	NSD—Norfolk	46	.00	L and S	NAS—Norfolk (Public Works)	23	.08
L	NSD—Norfolk (Portion)	39	.06	S	NSD—Oakland	41	.00

of this test were improved, the validity coefficients might be substantially higher.

## 5 READING COMPREHENSION TEST

Level	Activity	N	r	Level	Activity	N	rr
F and Q	NAS—Norfolk	39	.45	L	NAS—Norfolk	153	.12
Q	NAS—Norfolk	37	.31	L	NAS—Norfolk (AR)	129	.11
CQ and Q	NAS—Alameda	28	.43	L	NAS—Alameda (AR)	24	.35
Q	NSD—Oakland	20	.47	L	NSD—Oakland	47	.09
CQ and Q	NS—San Francisco	57	.02	L	NS—San Francisco	142	.06
Q	NS—Mare Island	68	.19	L	NS—Mare Island	171	.21
Q and L	NSD—Norfolk	60	.35	L	NS—Mare Island (Portion)	71	.28
L	NSD—Norfolk	46	.24	L and S	NAS—Norfolk (Public Works)	23	.13
L	NSD—Norfolk (Portion)	39	.32	S	NSD—Oakland	41	.31

This was the second-best test, next to the Supervisory Judgment Test, using as a rough index the percentage of groups on which a validity of .30 or more was obtained. Eight of the eighteen zero-order correlations are above .30 with the range between .47 and .02. An improvement in the reliability of the test from the present estimate of .649 might increase the validity. The same explanations would be appropriate here, as with

## 6 NUMERICAL REASONING TEST

Level	Activity	N	r	Level	Activity	N	r
F and Q	NAS—Norfolk	39	.19	L	NAS—Norfolk	153	.21
Q	NAS—Norfolk	37	.02	L	NAS—Norfolk (AR)	129	.21
CQ and Q	NAS—Alameda	28	.16	L	NAS—Alameda (AR)	24	.13
Q	NSD—Oakland	20	.21	L	NSD—Oakland	47	.05
CQ and Q	NS—San Francisco	57	.14	L	NS—San Francisco	142	.18
Q	NS—Mare Island	68	.10	L	NS—Mare Island	171	.24
Q and L	NSD—Norfolk	60	.25	L	NS—Mare Island (Portion)	71	.23
L	NSD—Norfolk	46	.05	L and S	NAS—Norfolk (Public Works)	23	.44
L	NSD—Norfolk (Portion)	39	.09	S	NSD—Oakland	41	.32

the Supervisory Judgment Test, in discussing the reasons for the differences in the obtained validity coefficients.

All of the correlations for this test are positive but low with only two exceptions. This test was somewhat better than the other tests for Leadingmen, but it was the poorest test for the levels above Leadingman.

### 7. BLUEPRINT READING TEST

<i>Level</i>	<i>Activity</i>	<i>N</i>	<i>r</i>	<i>Level</i>	<i>Activity</i>	<i>N</i>	<i>r</i>
F and Q	NAS - Norfolk	19	.19	L	NAS - Norfolk	153	.20
Q	NAS - Norfolk	17	.17	L	NAS - Norfolk (AR)	129	.23
CQ and Q	NAS - Alameda	28	.53	L	NAS - Alameda (AR)	24	-.01
Q	NSD - Oakland	20	.30	L	NSD - Oakland	47	-.03
CQ and Q	NS - San Francisco	57	.16	L	NS - San Francisco	142	.08
Q	NS - Mare Island	68	.12	L	NS - Mare Island	171	.12
Q and L	NSD - Norfolk	61	.20	L	NS - Mare Island (Porton)	71	.09
L	NSD - Norfolk	41	.05	1 and 5	NAS - Norfolk (Public Works)	23	-.07
L	NSD - Norfolk (Porton)	39	.10	5	NSD - Oakland	41	.23

The data for this test are inconclusive for two reasons: (1) Some of the samples included do not have to read blueprints in their work; and (2) Only ship blueprints were used. The correlations obtained in this study do not justify using this test for selection purposes in this program.

In addition to the correlations between job performance and test scores, the mean scores for each level on each test, by installation, were computed. These means tend to increase from the bottom supervisory level to the top.

The following table presents the intercorrelations of the tests, which are numbered in the same order as in the preceding tables, and estimates of the reliabilities of the tests, based on

*Test Intercorrelations and Reliabilities*

	2	1	4	5	6	7	<i>rit</i>
1.	.684	.542	.639	.658	.551	.506	.818
2.		.608	.533	.570	.612	.636	.816
3.			.397	.452	.452	.510	.878
4.				.531	.514	.418	.459
5.					.609	.426	.649
6.						.536	.854
7.							.614

Kuder-Richardson formula No. 21. These data are based on a total of 172 cases from three different Naval installations.

### *Conclusions*

The data presented seem to indicate that some of these tests can contribute effectively to the selection of supervisors in a wide variety of trades. When one takes into consideration the following factors, this conclusion is reinforced: (1) The correlations are based on original test content; (2) The ratings used as the criterion included cases where there were differences between the various raters; (3) The correlations were computed on groups in a variety of trades; and (4) The test reliabilities were in general low for individual prediction. Despite these factors, the data on the Supervisory Judgment and Reading Comprehension Tests indicate that these tests should improve the selection of supervisors. Both of these tests did relatively well at all supervisory levels and for various types of activities, while some of the other tests did better at one grade level than at another. Refinement of the administration of these tests should produce a better indication of the true value of these tests for supervisory selection. Tentative results from the validation of these tests on supervisors in white-collar occupations indicates that the Supervisory Judgment and Reading Comprehension Tests contribute to the selection of white-collar, as well as of blue-collar, supervisors.



# MEASURING THE TEACHING PERSONALITY

WALTER W. COOK  
University of Minnesota

AND

CARROLL H. LEEDS  
Mississippi State College for Women

THE need for measures of the personality factors related to success in teaching requires no elaboration. Many approaches to the problem are possible. One approach would first formulate a precise and analytical definition of the factors involved and then attempt to build an instrument which subjectively parallels the definition. Another would attempt to find among the personality measures now available those related to teaching success. Another, the one followed in this investigation, simply establishes two criterion groups of teachers, one having the desired personality characteristics to a high degree, the other to a low degree, and attempts to find factors which discriminate sharply between the two groups. An adequate attack involves elements of all three approaches. The difference is in emphasis. In accepting the third procedure the heavy theorizing is postponed until some measures are found which discriminate. The hypotheses upon which we operate remain highly tentative and no justification is sought other than successful discrimination.

In attacking the problem, we assumed that a teacher who is well liked by his pupils, and who establishes intimate and harmonious working relations with them has a high level of the personality traits we sought to measure. We believed that the emotionalized attitudes of pupils toward their teachers could be measured. We furthermore assumed that the attitudes of pupils toward teachers were the result of the teachers' attitudes toward pupils. Hence, if we could measure the attitudes of both teachers and pupils toward each other there should be a high relationship between them. Having demonstrated this

relationship, we could assume that the inventory of the teacher's attitudes toward pupils would be an effective measure of the teaching personality since it would indirectly measure the emotional responses which pupils make to teachers.

The validating procedures involved also the assumptions that a school principal who has worked with a teacher for some time can sense the emotional relationship between teacher and pupils and can discriminate accurately between those with extremely high and low rapport, and that an expert in the field of teacher-pupil relationships can visit a classroom and judge accurately the affective relationship which exists.

#### *Constructing the Teacher-Pupil Inventory*

The form of the teacher's-attitudes-toward-pupils Inventory was dictated by several considerations. It was assumed that such attitudes would constitute a very complex pattern, involving a great number of attitudes toward a great number of things and that therefore the simple linear concept of favorableness or unfavorableness of the Thurstone (7) and Remmers (4) techniques could not be applied. Furthermore, we wished to use criterion groups rather than the judgment of experts in the selection and weighting of the items. The Likert (3) technique of scoring was rejected because we could not assume that the attitudes of teachers toward each of the many aspects of child status and behavior would be normally distributed. In fact it seemed certain that most of them would be skewed. Furthermore, by using criterion groups, optimum scoring weights can be determined experimentally. The form of item recommended by Rosander (5), in which a situation is described and the subject is asked to choose among five possible responses, was rejected because it is difficult to build and time-consuming to administer—both of which tend to limit the sampling of attitudes. The chief argument for it is that it approximates more closely actual behavior but since the selection of the criterion groups is based on actual behavior this consideration was given less weight.

The item technique finally accepted for the Inventory makes each item essentially an attitude scale in itself. It was used by

Likert and called by Bird (2) the method of "Summated Ratings." It resembles the items of the *Strong's Interest Inventory* except that a five-point scale is used instead of a three-point. It was believed that three points would not discriminate adequately for our purpose. The directions with sample items were as follows:

### *Explanation and Directions*

There is considerable disagreement in educational circles relative to a variety of topics involving the relationship of teachers and pupils. These topics are still largely matters of opinion with little reinforcement from scientific evidence. It is the purpose of this Inventory to sample existing opinion in the area of teacher-pupil relationships. There are no right or wrong answers to the items of the Inventory. *Your own individual feeling about the matter is what is wanted. A frank and sincere response to each item is obviously necessary if this study is to serve its purpose. Your responses will have no effect whatever upon your personal or professional standing. Your name is not requested.* —

*Directions:* After reading the items carefully, indicate your response by encircling the number which most nearly represents your reaction to the topic in question. Think in terms of general situations rather than of individual cases. *Please respond to every item.*

The scale is as follows:

1. Strongly agree
2. Agree
3. Undecided or uncertain
4. Disagree
5. Strongly disagree

1. Most children are obedient	1	2	3	4	5
2. Shyness is preferable to boldness	1	2	3	4	5
3. Most pupils lack productive imagination	1	2	3	4	5
4. Children dress more sensibly nowadays	1	2	3	4	5
5. Children "should be seen and not heard"	1	2	3	4	5

### *Construction of Items*

In the construction of items five sources were canvassed, not with a view to uniqueness or independent measures but simply in order to secure a more adequate sampling of attitudes. These five areas were as follows:

1. *Moral status* of children in the opinion of adults, especially as concerns their adherence to adult-imposed standards, moral or otherwise.

Example: Pupils are naturally stubborn.

2. *Discipline* and problems of conduct in the classroom and elsewhere and methods employed in dealing with them.

Example: Pupils found writing notes should be severely punished.

3. *Principles of child development* and behavior related to ability, achievement, learning, motivation, and personality development.

Example: Success is more motivating than failure.

4. *Principles of education* related to philosophy, curriculum and administration.

Example: Pupils should be required to do more studying at home.

5. *Personal reactions of teacher*, his likes and dislikes, sources of irritation, etc.

Example: Without children life would be dull.

A total of 756 items were built for tryout purposes. These were placed in two experimental forms of the Inventory (A and B) with 378 items in each.

#### *Administration of Experimental Forms*

Instead of using experts to determine the relative importance and validity of the items and how various types of teachers ought to react to them, it was the plan of the investigation to determine these things by experimental procedure. This necessitated the employment of an outside criterion in selecting two groups of teachers. The criterion employed was the judgment of the principals of the schools in which the teachers taught.

Approximately 70 schools were visited by Dr. Leeds to seek the cooperation of the principals. Each principal was then asked to designate one or two teachers in his school whom the pupils liked very much, who had excellent working relations with them and who had the personality characteristics essential to effective teaching. He was then asked to designate one or two teachers who had these qualities in a very low degree. In some schools only one teacher was selected for testing, in others



two or three. All of the schools were public elementary, junior or senior high schools, tending toward the conservative type, in the states of Ohio and Pennsylvania. Table 1 gives more detailed information about the schools and the number of teachers tested in each.

Each designated teacher was then visited and asked to cooperate in a research project by filling out Form A of the Inventory. The teacher was assured that no one but the investigator would examine the completed Inventory. A few weeks later Dr. Leeds visited the teacher again, picked up the completed Inventory and left Form B to be filled out and mailed. This procedure was continued until both Forms A and B of the Inventory had been completed by 100 superior and 100 inferior teachers.

TABLE 1

*Number of Schools and Teachers Representing Different-Sized Communities and Different School Levels from Which Two Groups of 100 Teachers Each Were Drawn*

Size of Community		Superior Group		Inferior Group	
		No. of Schools	No. of Teachers	No. of Schools	No. of Teachers
Cities with population of 100,000 and over	Sr. High	2	3	1	4
	Jr. High	1	1	1	2
	Elem	9	14	8	16
Cities with population between 1,000 and 100,000	Sr. High	9	21	8	20
	Jr. High	5	10	4	9
	Elem	17	30	18	33
County schools and towns with population of 1,000 and under	High	14*	9	11*	9
	Elem		12		7
Total		57	100	51	100

\* Both elementary and secondary grades were housed in one building

### *The Selection of Items for the Final Inventory*

The statistic, chi-square, was used to determine the extent to which each item differentiated between the two groups. It was found that 115 of the 756 items discriminated at the 5 per cent level. However, it was necessary to take factors other than discriminating power into consideration in the selection of the 164 items included in the final Inventory. Some of the highly discriminating items duplicated others in content,

some were poorly stated and others showed a response pattern that was difficult to interpret.

### *Weighted Scoring Procedure*

An inspection of the responses of the criterion groups to the items selected for inclusion in the Inventory revealed that most of them were far from normal in character. Many were bimodal. Others tended to pile up at one or the other of the extremes. It was obvious that the simplified scoring procedure of Likert could not be employed.

It was decided to use the Kelley (6) formula for weighted scores developed for use with the *Strong's Interest Inventory*. This formula weights each of the five possible responses on a nine-point scale from minus four to plus four, the weighting being determined by the differences in the responses of the two criterion groups. Two items are given below with the percentage of teachers in each group making each response and the weights determined by the formula. It is obvious that such weightings could not be determined except by empirical methods.

*Item: Activities in the modern school take too much time from classwork.*

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
	1	2	3	4	5
Superior teachers . . . . .	4	19	20	47	10
Inferior teachers . . . . .	6	28	19	41	6
Difference . . . . .	-2	-9	1	6	4
Scoring weight . . . . .	-1	-2	0	2	1

*Item: It is sometimes necessary to break promises made to children.*

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
	1	2	3	4	5
Superior teachers . . . . .	2	38	12	44	4
Inferior teachers . . . . .	1	53	15	22	9
Difference . . . . .	1	-15	-3	22	-5
Scoring weight . . . . .	0	-4	-1	4	-1

*Simplified Scoring Procedure*

The system of scoring described above necessitates the cumbersome process of adding algebraically a long list of positive and negative numbers. Since most studies of weighting procedure indicate that a simpler process does not greatly influence the validity of the test, such a process was tried out. The simplified scoring scheme adopted assigns a plus-one weight to each response with a positive weight and disregards all responses with negative or zero weights. The correlation between the weighted and simplified scores for the inventory was .97. Validity coefficients will be presented later for both systems of scoring.

*Measuring the Validity and Reliability of the Teacher-Pupil Inventory*

With the discriminating items selected and scoring weights established for each of the five possible responses to each item, the next step was to measure the validity of the Inventory. This was done by administering the Inventory to a new, and in this case, random group of 100 teachers in grades 4-6 inclusive. Three outside criteria of teaching personality were established for this group and correlations were computed between the scores on the three criteria and the scores on the Inventory.

The three outside criteria were established by the following procedures. One criterion involved the rating of the teacher by his pupils. The intermediate grades were selected for the try-out because it was felt that pupils at this level would be sufficiently mature to respond to simple items in a questionnaire and that they would be less biased in giving frank answers than the more sophisticated high-school pupils. Ratings were obtained by at least 25 pupils on each of the 100 teachers.

The Pupils' Rating Scale consisted of 50 simple *Yes-No-?* items of the following type:

- |  |     |    |   |
|--|-----|----|---|
| 1. Do you like school?                               | Yes | No | ? |
| 2. Does the teacher make the schoolwork interesting? | Yes | No | ? |
| 3. Is the teacher often "bossy"?                     | Yes | No | ? |
| 4. Does the teacher force her ideas on the pupils?   | Yes | No | ? |

- |  |     |    |   |
|--|-----|----|---|
| 5. Is the teacher usually kind to you? | Yes | No | } |
| 6. Do most pupils like this teacher?   | Yes | No | } |

The Pupils' Rating Scale was scored by giving a plus-one (+1) weight to each favorable response and a minus-one (-1) to each unfavorable. Space was provided for each pupil to write his answers to the following questions: (1) Why I like this teacher, (2) Why I don't like this teacher, and (3) What grade or mark do you usually receive from the teacher? The Pupils' Scale was administered by Dr. Leeds with no one else in the room but the pupils. Assurance was given to the pupils that no one other than Dr. Leeds would see the ratings. The ratings were not signed.

The reliability of the pupils' ratings was determined by two procedures. First, 200 pupil ratings were chosen at random and their scores on the odd items correlated with the even items. This procedure gave a reliability of .94. The second procedure involved selecting two groups of ten ratings each for each of the 100 teachers, correlating the average scores for the two groups and then stepping up the correlation by the Spearman-Brown formula to equal that of 25 ratings. This method gave a reliability of .93 for 25 ratings.

Previous to the establishment of the pupil-rating criterion two other criteria had been established. This was to prevent the pupil's criterion from influencing in any way the other two. One of these was a rating by the principal of each school, the other a rating by Dr. Leeds based on a visit to each teacher's classroom. The Principal's Rating Scale was designed specifically for this purpose and required the marking of descriptive ratings on the following traits:

1. Disciplinary ability
2. "Personnel" vs. "subject-matter" point of view
3. Attitude toward children
4. Understanding of pupil behavior problems
5. Personality adjustment
6. Attitude of pupils toward the teacher

The possible score range on this Rating Scale was from -14 to +14. The actual ratings ranged from -9 to +14. The re-

liability of the Scale as determined by the split-half method was .87.

The Scale used by Dr. Leeds in rating the 100 teachers was a modification of Baxter's Rating Scale of the *Teacher's Personal Effectiveness* (1). It was designed for the purpose of evaluating a teacher's personal effectiveness in the classroom by noting and analyzing pupils' responses to the teacher's conduct. For this study, the Scale was modified by omitting the responses of the pupils, omitting two irrelevant items and adding an item relating to the teacher's disciplinary ability.

Enough time was spent in the classroom of each teacher to observe his effectiveness in:

1. Maintaining discipline
2. Creating a friendly classroom atmosphere
3. Establishing a feeling of security
4. Exerting a stabilizing influence on the class
5. Developing pupil self-reliance

The possible score range on the Teacher's Personal Effectiveness Scale was from +10 to -10. The actual range was from +10 to -9. The reliability of the Scale as determined by the split-half method was .92. The high reliability of this Scale and the Principal's Scale could be explained by the tendency for a judge to give similar rating in traits that seem to him logically related.

The correlations between the *Teacher-Pupil Inventory* scores (weighted and simplified) and the three criterion ratings by pupils, principals, and expert are shown in Table 2. When the scores of the three criteria are combined, with equal weight given to each by standard score procedure, the validity coefficients of the Inventory are .59 and .60 for the weighted and simplified scores respectively. When the three criteria are combined with multiple weighting the validity coefficient is not significantly different (.60). The single criterion measure which has the highest relationship with the *Teacher-Pupil Inventory* is the expert's rating. The correlation is .49. The next highest is the pupil's rating, .45, and the lowest is the principal's rating, .43.

TABLE 2  
*Correlations between Teacher-Pupil Inventory Scores (Weighted and Simplified)  
 and Three Criterion Ratings by Pupils, Principal and Expert  
 for an Unselected Group of 100 Teachers\**

	Weighted Scoring	Simplified Scoring
Inventory vs. Pupil's Ratings .....	.45	.46
Inventory vs. Principal's Ratings .....	.43	.45
Inventory vs. Expert's Ratings .....	.49	.49
Inventory vs. Three Validating Criteria Combined, Equal Weights .....	.59	.60
Inventory vs. Three Validating Criteria Combined, Multiple Weights .....	.60	..
Inventory vs. Combination of Pupil's and Expert's Ratings .....	.58	.58
Inventory vs. Combination of Pupil's and Principal's Rating .....	.53	.54
Inventory vs. Combination of Expert's and Principal's Rating .....	.54	.54
Expert's Rating vs. Principal's Rating .....	.48	..
Expert's Ratings vs. Pupil's Rating .....	.33	..
Principal's Rating vs. Pupil's Rating .....	.39	..

\* All correlations are significant at the 1 per cent level.

The reliability of the *Teacher-Pupil Inventory* as determined by the split-half method using the simplified scoring pro-

TABLE 3  
*Distributions of Teacher-Pupil Inventory Scores for the Superior, Unselected,  
 and Inferior Groups of Teachers*

Scores	Frequencies		
	Superior	Unselected	Inferior
301- 340	5		
261- 300	8		
221- 260	11	5	
181- 220	12	13	4
141- 180	9	11	6
101- 140	14	16	5
61- 100	18	18	4
21- 60	6	9	13
- 19- 20	10	9	11
- 59-(- 20)	5	8	17
- 99-(- 60)	2	6	11
- 139-(- 100)		2	13
- 179-(- 140)		3	8
- 219-(- 180)			2
- 259-(- 220)			5
- 299-(- 260)			1
<i>N</i> .....	100	100	100
Median .....	126.2	89.4	- 36.0
Mean .....	133.2	77.6	- 40.8
S.D. ....	102.8	100.2	112.8

cedure with the 100 randomly selected teachers at the 4th, 5th, and 6th grade levels was .89.

The distributions of the scores on the *Teacher-Pupil Inventory* for the superior, unselected and inferior groups of teachers are shown in Table 3. None of the distributions deviate significantly from the normal at the 5 per cent level. The unselected group of teachers proved to be a rather high scoring group. This could be anticipated since the intermediate teachers in the two other groups were superior to the junior and senior high-school teachers. The primary teachers tended to rate the highest. However, the highest scoring teacher in all of the testing was a high-school counselor of girls.

### Summary

1. When we assume that by "teaching personality" we mean those characteristics of the teacher's behavior related to the emotional responses of pupils and the ability to establish intimate and harmonious working relations with them, we find that "teaching personality" can be measured.

2. We find that the attitude of individual teachers toward pupils is significantly related to the pupils' attitudes toward the teachers. Both sets of attitudes can be measured with a reliability approaching .90 and the relationship between them established with a correlation as high as .46.

3. Pupils' ratings of teachers at the intermediate-grade levels are reliable and valid. There is a significant relationship between their ratings and those of the principal and of an expert. Pupils' ratings of teachers correlate with principals' ratings, .39, and with an expert's ratings, .33. The expert's and the principals' ratings correlate .48.

4. The "teaching personality" can be measured with as high a validity as can academic aptitude, the correlation with three criteria combined being .60.

### REFERENCES

1. Baxter, Bernice. "Rating Teachers' Personal Effectiveness." *National Education Association Journal*, XXVII (1938), p. 81.

2. Bird, Charles. *Social Psychology*. New York: D. Appleton-Century Company, 1940, p. 159.
3. Likert, R. A. "A Technique for the Measurement of Attitudes." *Archives of Psychology*, XXII (1932), 5-55.
4. Remmers, H. H. "Generalized Attitude Scales." *Journal of Social Psychology*, V (1934), 296-312.
5. Rosander, A. C. "Attitude Scale Based upon Behaviour Situations." *Journal of Social Psychology*, VIII (1937), 3-15.
6. Strong, Jr., E. K. *Vocational Interests of Men and Women*. Stanford University: Stanford University Press, 1943, chaps. xxiii, xxiv.
7. Thurstone, L. L. and Chave, E. J. *The Measurement of Attitude*. Chicago: Chicago University Press, 1929.



# THE INSIGHT OF COLLEGE STUDENTS INTO THE ITEMS ON A PERSONALITY TEST

J. A. MORRIS KIMBER  
University of Southern California

ITEMS on the self-inventory type of personality test have been recognized to be different from those on other types of tests in one important particular. With the individual motivated to secure a favorable score or to make a certain impression, he probably has greater ability to discern the answer which would contribute toward attaining his end than he has on a test of achievement, intelligence, or aptitude. This ability is closely connected with his insight into the items which appear on the inventory. His intention may not be accomplished unless he has this needed insight.

Superficial consideration of the fact that scores on personality tests may be influenced by the desire to secure favorable scores or to create certain impressions may lead to a rejection of such tests as instruments of scientific value. Further thought, however, has led some psychologists to propose inquiries into the possibility of controlling this influenceability or at least of measuring it. Some have even suggested that this influenceability may be an asset, and that a study of the subject's score as influenced by his desire to make a certain impression or resulting from some "unorthodox" procedure may render possible a diagnosis of the personality of the subject.

To what extent do individuals have insight into the "right" responses to the items on a typical personality inventory? An important correlate is the question as to whether insight is identical with or closely related to general intelligence. The present study, conducted with college students, is an attempt to answer these questions.

*Personality Test Selected.*—The California Test of Person-

ality, Secondary Form A, was selected for this study. It was believed advisable to use a test which in its standardization had furnished the same norms for men and women, since both men and women were included in the group to be studied. The selection of the Secondary Series was made on the basis of the suitability of the questions. Since a few of the items on the Adult Form deal with employment experience and with the exercise of the franchise, it was believed that in spite of the fact that the Secondary Form appears to be designed for high-school rather than for college students its use was preferable.

*Intelligence Test Selected.*—Guilford's First Nebraska Revision of *The Army Alpha Examination* was selected not only because it is still regarded as one of the best group tests of intelligence but because of the fact that factorial analysis had indicated that three factors, *V*, *N*, and *R* could be scored separately.<sup>1</sup> Another important reason for using the test in this particular study is the fact that here, too, the published norms for men and women are the same.

*Procedure.*—The procedure in carrying out the study required the administration of the *California Test of Personality* twice to students registered in General Studies 51, the introductory course in Psychology at the University of Southern California. Later the Army Alpha was given to the same group.

*First Administration of the California Test of Personality (Test 1).*—Following a few words concerning the value of trying to understand people, the investigator addressed the class as follows:

Today we are giving you a copy of a personality inventory. The object we have in giving it is to find out how you interpret the reactions of people. Read each question carefully. Then answer the question the way you think a happy and well-adjusted student at U. S. C. would answer it. Remember, we are not asking you to tell the way you are. We are asking you to answer the questions the way you think a happy and well-adjusted student at U. S. C. would answer them.

*Second Administration of the California Test of Personality*

---

<sup>1</sup> Guilford, J. P. "A New Revision of the Army Alpha Examination and a Weighted Scoring for the Three Primary Factors." *Journal of Applied Psychology*, XXII (1938), 239-246.

(Test 2).—The procedure here was the standard one for the administration of this test.

*Administration of The Army Alpha Examination—First Nebraska Revision.*—This test was administered according to the "Directions for Giving" which appear in the *Manual* for the First Nebraska Revision of the *Army Alpha Examination*.

*Personnel of the Group Tested.*—Of the 389 students who took Test 1 (First Administration of the *California Test of Personality*), Test 2 (Second Administration of the *California Test of Personality*), and the Army Alpha, 238 (61 per cent) were women, and 151 (39 per cent) were men. Of the 151 men tested, 92 (61 per cent) were trainees in the Navy program operating at the University, while 59 (39 per cent) were civilians. All of the women who took the tests were civilians and were assumed to constitute a representative group of university women. It is believed that in the case of trainees there was a greater selectivity as a result of Naval policies. In the interests of securing homogeneity, the scores of students over twenty-four years of age were not included.

*Reliability of the Instrument Used.*—Owing to the fact that the standardization of the *California Test of Personality*, Secondary Form A, had been carried out with secondary-school students, it was regarded as important to determine its reliability as administered to college students. Further, since Test 1 imposed entirely new conditions, it was essential to determine the degree of reliability which the test had when given under these conditions. Spearman-Brown corrected coefficients of correlation between even and odd numbered items on the two administrations of the test were as follows:

Test 1		Test 2	
167 men .....	.936	167 men .....	.947
259 women .....	.907	259 women .....	.902

It was of interest to find that these correlations demonstrate a reliability which is even greater than that claimed for the test by its writers.

Scoring of the tests revealed significant differences between

mean total scores on Test 1 and Test 2 in the case of both men and women. While the differences between the scores made by men and women are discussed subsequently, it is of interest to note that mean total scores of men and of women were almost exactly the same on Test 2, the second administration of the *California Test of Personality*, while on Test 1, the first administration of the test, there was a substantial difference (C.R. 2.6).

It might easily be assumed that since the differences between the mean total scores on Test 1 and Test 2 were significant in the case of both men and women, there would be almost unanimity with regard to students receiving lower scores on Test 2. This was not found to be true. A substantial number of both men and women received higher scores on Test 2; a larger number of men than of women received the higher scores on Test 2.

*The Investigation as a Reification of Insight.*—It is the conclusion that the significant differences between the results ob-

TABLE 1  
*Total Scores on the First and Second Administration of The California Test of Personality*

	Test 1	Test 2	Critical ratio
Men ( $N=151$ )			
Mean .....	158.8	153.7	4.6
Standard deviation .....	14.39	13.26	1.4
Women ( $N=238$ )			
Mean .....	162.5	153.9	10.8
Standard deviation .....	11.94	13.40	2.6

TABLE 2  
*Comparison of Students in Terms of the Relation of Their Scores on Test 1 and Test 2*

	Men		Women	
	Number	Per cent	Number	Per cent
Students who received higher score on Test 1 than on Test 2 .....	99	59.28	193	74.52
Students who received lower score on Test 1 than on Test 2 .....	63	37.72	59	22.78
Students who received the same score on Test 1 as on Test 2 .....	5	3.00	7	2.70
	167	100.00	259	100.00

tained on Test 1 and those obtained on Test 2 for both men and women may be attributed to an insight into the meaning of the items. Insight is regarded as demonstrated when a student sees into the meaning of an item and when instructed not to answer for himself but with another hypothetical person in mind—in this case a happy and well-adjusted student at U. S. C.—responds to that item in agreement with the key furnished by the authors of the inventory, the "right" answers on the key presumably being the responses of the majority of well-adjusted persons.

It would appear that the most convincing demonstration of such a concept as insight would be the measurement of such insight. Thus it is desired to show that this study constitutes a step in the measurement of insight.

*The Measurement of Insight.*—Two possible methods are available for determining the degree of insight which students had into the items of the *California Test of Personality* as administered experimentally under the two conditions described

TABLE 3  
*Number of Items Answered According to Key in Terms of Comparison  
between Scores on Test 1 and Test 2*

	Men ( $N=167$ )		Women ( $N=259$ )	
	Number	Per cent	Number	Per cent
Items answered with a higher percentage of students responding according to key on Test 1 than on Test 2 .....	144	80.0	133	73.9
Items answered with a lower percentage of students responding according to key on Test 1 than on Test 2 .....	21	11.7	22	12.2
Items answered with the same percentage of students responding according to key on Test 1 as on Test 2 .....	15	8.3	25	13.9
	180	100.0	180	100.0

TABLE 4  
*Coefficients of Correlation between Mean Total Scores on Test 1 and Test 2*

	$r$	P.E.
Men ( $N=151$ ) .....	.521	.04
Women ( $N=238$ ) .....	.535	.03

above. The first of these is the study of total scores, while the second is the study of the frequency of responses to the items. The results of the first method have been presented in Table 1. The results of the second method appear in summarized form in Table 3.

Numerous explanations might be given of the difference in scores between Test 1 and Test 2 for individual students. What would appear to be almost certain evidence of a high degree of insight would be exemplified by the student who made a high score on Test 1 but a low score on Test 2. What would appear to be definite evidence of a low degree of insight would be exemplified by the student who made a low score on Test 1 but a high score on Test 2. However, whatever the score on Test 2, it is believed that a high score on Test 1 indicates a high degree of insight into the items whereas a low score on Test 1 indicates a low degree of insight into these items.

*Insight and Personality.*—It will be seen that there is a relatively high correlation for both men and women between the scores of students on the *California Test of Personality* administered as Test 1 and the same test administered as Test 2. These correlations appear in Table 4. It will also be observed in Table 1 that the standard deviation of Test 1 is approximately the same as that for Test 2. The question may thus be raised: Do we have in Test 1 a good disguised personality test?

It is clear from the coefficients of correlation that whatever is being measured by Test 2 is to some extent being measured by Test 1, and vice versa. Not until objective validation can be secured can it be demonstrated that the test given as Test 1 is as good as or better than the test given as Test 2. It is the hypothesis, however, that the test administered as Test 1 may be a good measure of personality. Here the student is using something of the projection technique. He finds himself free from reproach and blame when he answers the questions. It is not his concern whether a happy and well-adjusted student suffers "often from annoying eyestrain," for example. He does not need to protect this happy and well-adjusted student from criticism because he finds that he is "tired a great deal of the time." Hence the hypothesis is presented that the answers on Test 1

may be more accurate since they are more objective. Test 1 is not invalidated by the student's having insight into the items. The insight is taken into consideration and the student's score is the score on a personality test given under prescribed conditions for a particular purpose. It seems clear that either Test 1 or Test 2 may be regarded as a personality test but that Test 1 is a better disguised personality test.

*Personality and Intelligence.*—From numerous studies it has been found that there is a relatively low correlation between measures of personality and measures of intelligence. Findings of the present study confirm those formerly made and also indicate that correlations between personality and the various factors established by Guilford are also low. These results appear in Table 5.

TABLE 5

*Correlations between Test 2 (Second Administration of the California Test of Personality) and the Army Alpha Examination*

	Men (N = 151)		Women (N = 238)	
	r	P.E.	r	P.E.
Test 2 and Army Alpha total .....	-.041	.05	.106	.04
Test 2 and Army Alpha, "V" factor (Verbal) ...	-.113	.05	.087	.04
Test 2 and Army Alpha, "N" factor (Numerical) .	.005	.05	.110	.04
Test 2 and Army Alpha, "R" factor (Relational) .	-.176	.05	.046	.04

TABLE 6

*Correlation between Test 1 (First Administration of the California Test of Personality) and the Army Alpha Examination*

	Men (N = 151)		Women (N = 238)	
	r	P.E.	r	P.E.
Test 1 and Army Alpha total .....	.102	.05	.211	.04
Test 1 and Army Alpha, "V" factor (Verbal) ....	.130	.05	.196	.04
Test 1 and Army Alpha, "N" factor (Numerical) .	.063	.05	.207	.04
Test 1 and Army Alpha, "R" factor (Relational) .	.123	.05	.137	.04

*Insight and Intelligence.*—The question naturally arises as to whether the insight shown by students in this study is largely a question of general intelligence. Correlations between the scores on Test 1 and the Army Alpha appear in Table 6. The

low correlations indicate that the highly intelligent student is not necessarily the one who possesses a high degree of insight, nor is the highly "insightful" student necessarily the one of highest intelligence.

The low correlation between scores on Test 1 with scores on the *V* factor is of interest in as much as insight has sometimes been thought of as closely related to verbal or linguistic ability.

*Sex Differences.*—One of the findings of this investigation has been that certain differences exist between the responses of men and women on the tests administered. As was formerly stated, one reason for the selection of the two tests used, viz., the *Cal-*

TABLE 7

*A Comparison of Scores of Men and Women in Terms of Means, Standard Deviations, and Percentages who Responded According to the Key on the Tests Administered*

	Men	Women	C R.
Mean, Test 1 .....	158.8	162.5	2.6
S.D., Test 1 .....	14.39	11.94	2.5
Mean, Test 2 .....	153.7	153.9	.1
S.D., Test 2 .....	13.26	13.40	.0
Mean, Army Alpha total .....	161.8	149.5	5.2
S.D., Army Alpha total .....	25.14	18.15	4.2
Mean, Army Alpha, "V" factor .....	133.1	127.2	3.3
S.D., Army Alpha, "V" factor .....	14.36	20.40	4.8
Mean, Army Alpha, "N" factor .....	53.8	42.5	12.7
S.D., Army Alpha, "N" factor .....	8.86	8.09	1.2
Mean, Army Alpha, "R" factor .....	44.2	43.2	1.6
S.D., Army Alpha, "R" factor .....	5.80	6.09	.7
Percentage who made higher score on Test 1 .....	59	74	3.2
Percentage who made same score on Tests 1 and 2 .....	3	3	.0
Percentage who made lower score on Test 1 .....	38	23	3.3

*formia Test of Personality*, Secondary Form A, and the *Army Alpha Examination*, First Nebraska Revision, was the fact that the published norms are the same for men and for women.

The results given in Table 7 indicate that the *California Test of Personality* given as Test 2, and thus closely resembling conditions under which the test was standardized, confirmed the authors' findings that there was no significant difference between the scores of men and women. When, however, the test was given as Test 1, presumably as a test of insight into the responses of a hypothetical "happy and well-adjusted student," the results are in marked contrast. Here there appear relatively



high scores for women and relatively low scores for men, the difference between the means being 3.7. (C.R. 2.6.)

This difference between the mean scores of men and of women on Test 1 is even more striking when it is noted that while there is a very low correlation between the scores of both men and women with the scores on the *Army Alpha Examination* or any of its factors, men had significantly higher scores for the total, for the *V* factor, and for the *N* factor. The men, therefore, although demonstrated to be decidedly superior to the women in intelligence are decidedly inferior to the women in insight, assuming that Test 1 is a measure of insight.

When the further fact is considered that the men differed from the women in that a significantly higher proportion of men secured higher scores on Test 2 than on Test 1, the conclusion seems justified that college women differ strikingly from college men in the possession of a greater degree of insight.

#### *Summary and Conclusions*

The *California Test of Personality* was twice administered to approximately four hundred students registered in a beginning course in psychology at The University of Southern California. On the first occasion students were instructed to answer the questions as they believed that a happy and well-adjusted student at Southern California would answer them. On the second occasion students answered for themselves. The *Army Alpha Examination*, a test of intelligence, was later administered to the group.

*Conclusions.*—Among the findings which indicate that students have a high degree of insight into the items on the test were (1) the high scores on the first test and (2) the amount of difference between mean total scores on the first and second tests, the higher score appearing on the first test for both men and women.

Students differ greatly in the amount of insight which they possess. Some students secured high scores on the first test and low scores on the second. A high degree of insight is regarded as responsible for this difference. Some students, on the other hand, secured low scores on the first test and high scores on the

second. A low degree of insight is regarded as responsible for this difference.

Women exceed men in the amount of insight shown, as judged by their having higher scores on the first test. Where the test was given the second time, there was little difference between the scores of men and of women. Thus the men and women examined appear to resemble each other in personality adjustment but to differ in their insight into items on a typical personality test.

There is a relatively high degree of correlation between the scores on the first test and scores on the second test for both men and women. This fact supports the view that not only can the first test be regarded as a measure of personality or of some personality trait but can also be said to measure to some considerable extent whatever is being measured when the test is administered under standard conditions.

Correlations between scores on the second test with scores on *The Army Alpha Examination* were low. This fact supports previous studies which have demonstrated little relationship to exist between personality adjustment and intelligence.

Correlations between scores on the first test with scores on *The Army Alpha Examination* were also low. This fact is an indication that there is little relationship between intelligence and insight. Further support is given to this conclusion when it is noted that the men, though superior in intelligence, were inferior in insight.

### TESTS

*California Test of Personality*—Secondary Series, A Profile of Personal and Social Adjustment. Ernest W. Tiegs, Willis W. Clark, and Louis P. Thorpe. Los Angeles: California Test Bureau, 1939.

*Army Alpha Examination*—First Nebraska Revision. J. P. Guilford. Beverly Hills, California: Sheridan Supply Company, 1938.

# A STUDY IN TOOL USAGE

DAVID L. EKBERG  
The Standard Register Company

## *Background*

WHILE setting up a tentative battery of tests in 1941 for the selection of trainees for a specialized operation (called Operation "A") for the Standard Register Company the writer was fortunate in working with a foreman, Frank Voelkl, who, in addition to being well trained in the operation with which we were concerned, was also a great deal of help in analyzing the abilities required for an adequate performance of the duties.

This foreman mentioned one day that the actions of some men were characterized by purposeless motions which slowed their work. He related a common case of the manner in which one man would set punches and perforators on a machine by first using a tool to loosen a whole row of nuts, then unscrewing the entire row, using both hands. The slower man would loosen one nut at a time and then unscrew that nut with one hand before going on to the next nut. Obviously, the first man could, in the course of the day, accomplish more work of this type.

## *Analysis*

On the surface, this problem appeared to be one which could be corrected by the proper training. However, a study of individual cases revealed that training had been of little aid in correcting the problem. There appeared to be a knowledge of and/or an ability for speed and motion-saving which some men brought to the job with them. The diagnosis of this ability was considered a worth-while testing objective.

## *Criterion*

For practical reasons a rating of the subjects on the quality of time-saving or motion-saving was out of the question. How-

ever, it was possible to rank the subjects on the basis of general effectiveness on their job. This criterion was settled upon as being useful, though not ideal. Three qualified raters, all foremen, ranked the subjects. These rankings were pooled and used as our criterion.

### *Relationship with General Intelligence*

Subjectively, it would appear that a measure of intelligence should be useful in predicting success on this job if only because of the element of motion-saving which has been discussed. However, as the following table indicates, this assumption was disproved:

TABLE 1  
*Correlation, Foremen Rankings on General Job Effectiveness with Bengé Basic Employment Test—Two Groups Mechanical Workers*

Group	Number of Cases	Rho	Probable Error
Operators "A" .....	40	.23	± .11
Operators "B" .....	36	.08	....

### *New Measure Devised*

In the attempt to measure the ability in which we were interested the writer devised a performance test, called *Screw-board #1*.<sup>2</sup> The nature of the device is indicated in the photograph (Figure I). Pains were taken to case-harden the steel parts in order that the test would not wear with use, creating a variation in the standardization results. The bolts and screws were also carefully ground in to allow each one to run smoothly.

No attempt was made entirely to eliminate two-hand coordination, finger dexterity, or general intelligence, but rather it was intended that the test should duplicate, as a partial work-sample, a portion of the job. The directions were planned, however, to discover insight into a tool-usage problem. The subject was not permitted to see the test until immediately prior to the administration.

<sup>2</sup> Mr. Eugene J. Bengé was of invaluable assistance in the development of this measure as well as of all other tests and batteries used by the firm concerned.

*Directions*

Set up board and tools in standardized manner. Say:  
"The idea here is to remove the screws and bolts from the plate and put them in the pan (point) as quickly as you can. Use the tools as necessary. You may use both hands. Work as fast as you can. Now begin."

Note time and say:

"Now I want you to put them back in the plate and tighten them down with your fingers. Begin."

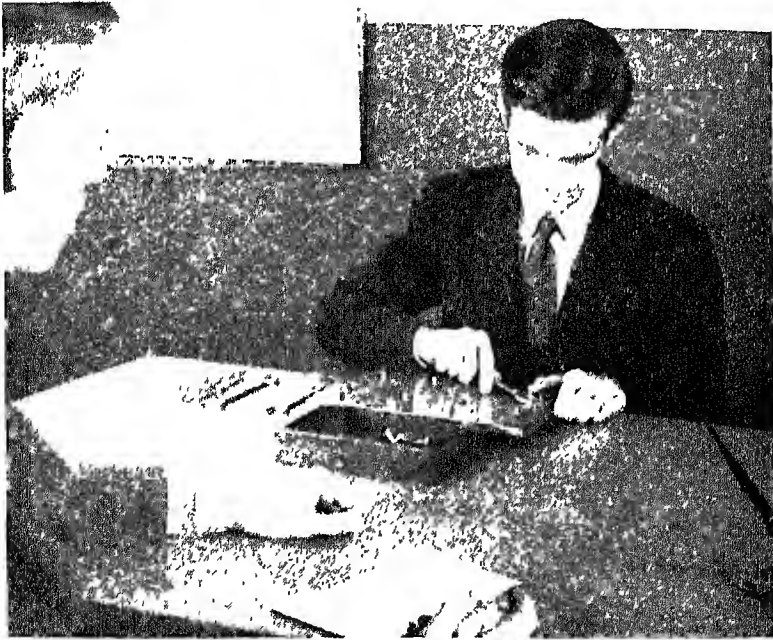


Figure I

Note time and tighten bolts with tools. Bolts should not be tight enough to move the board when using tools but, rather, just tight enough to prevent unscrewing them with the fingers alone. Then say:

"Now I want you to take the bolts out again the same way you did before. See how much faster you can work. Begin."

Note time and say:

"All right, put the bolts back in, tightening them down with your hands. Begin."

Note time. Add four sub-scores.

No practice period is given and no specific directions in the method of operation are suggested.

### Reliability

The present test, consisting of two trials in which the bolts are removed and two in which they are replaced, has a split-half, Spearman-Brown reliability coefficient of .89, based on 396 heterogeneous cases. The standard error of measurement is .495 minutes.

### Validity

The Screwboard was administered experimentally first to 40 Operators "A" along with a number of other tests. It was later administered to other similar machine operators. The validity results are shown in Table 2.

TABLE 2

*Correlation, Foremen Rankings on General Job Tests, and on Screwboard #1  
Four Groups of Mechanical Workers in a Certain Manufacturing  
Business Firm and Systems*

Group	Number of Cases	Rho	Comparable Pearson's	Probable Error
Operators "A" . . . .	40	.49	51	± .083
Operators "B" . . . .	36	.26	27	± .11
Operators "C" . . . .	15	.13	13	± .18
Operators "D" . . . .	9	.68	70	± .127

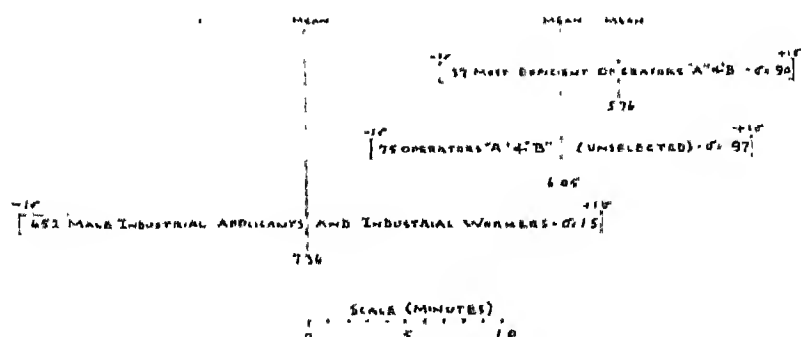


Figure II. Validation Analysis- Screwboard #1.

Showing relationship between test scores of general adult workers and scores of skilled operators in a certain manufacturing business firms and systems. None of the operators shown was selected by means of any test. The most efficient operators have the highest average score and the narrowest range of scores

Due to the homogeneity of the above groups the rank order correlations presented do not give a true picture of the value of the test. As a true indication of the validity of the test, the mean score of the group of Operators "A" and "B" combined, is 6.05 minutes or .87 sigma above the mean score of 652 male industrial applicants. The mean score of the 37 most efficient men picked from this group is 1.07 sigma above the mean of the applicant group. This analysis is pictured graphically in Figure 11. Furthermore, the battery of which this test is one portion, is accepted by the supervisors of the department concerned as a necessary selection tool. The results have been adequately proved in five years of operation.

#### *Correlation with Other Measures*

The highest correlation with any other measure thus far attempted is a test of two-hand coordination. These relationships are indicated in Table 3.

TABLE 3  
*Correlation Scorecard #1 with Other Measures*

Test	Type of Test	Number of Cases	$r$	P.E.
Benge Disc	Two-Hand Coordination	40	.60 ± .071	
Benge Basic Employment	General Intelligence	625	.30 ± .025	
Assembly #1 (SR Co.)	Mechanical Assembly	40	.43 ± .091	
Minn. Paper Formboard AA	Spatial Relations	63	.23 ± .038	

#### *Relationship with Age*

Using 200 random male cases a Pearson correlation between test scores and age at last birthday revealed a relationship of .33 ± .043 P.E. There appears to be a tendency for older men to score higher on this test than do younger men. It is probable that age, in this case, is related to experience with tools and mechanical work-habits but lacking an adequate standard of mechanical experience this study must await future work.

#### *Nature of Distribution*

This test, used on our populations, is characterized by skewed distributions of moderate degree. Using the formula:

$$Sk = \frac{3(\text{Mean} - \text{Median})}{\text{Sigma}}$$

TABLE 4

*Percentile Ranks of Scores on the #1 Balance, 1770, and #2 Industrial Mechanical Work and Industrial Worker*

Time in Minutes and Seconds	Male N = 62		Female N = 46		Continuation		
	Mean	7.36	Mean	8.87	Time	Male	Female
	Median	6.87	Median	8.12			
	Sigma	1.50	Sigma	1.89			
4.4	1		..		9.3	92	72
4.5	2		1		9.4	92	74
4.6	2		1		9.5	93	75
4.7	2		1		9.6	93	76
4.8	3		1		9.7	94	77
4.9	3		1		9.8	95	78
5.0	4		1		9.9	95	79
5.1	5		1		10.0	95	80
5.2	6		1		10.1	96	80
5.3	7		1		10.2	96	81
5.4	9		1		10.3	97	81
5.5	11		1		10.4	97	82
5.6	13		1		10.5	97	82
5.7	15		1		10.6	97	83
5.8	18		1		10.7	98	83
5.9	21		2		10.8	98	84
6.0	24		3		10.9	98	85
6.1	28		4		11.0	98	86
6.2	31		6		11.1	98	87
6.3	34		7		11.2	99	88
6.4	37		8		11.3	99	89
6.5	40		9		11.4	99	90
6.6	43		10		11.5	99	91
6.7	46		11		11.6	99	91
6.8	49		13		11.7	99	92
6.9	51		15		11.8	99	93
7.0	53		17		11.9	99	93
7.1	56		19		12.0	99	94
7.2	58		21		12.1	99	94
7.3	61		24		12.2	99	95
7.4	63		27		12.3	99	95
7.5	66		31		12.4	..	96
7.6	68		35		12.5	..	96
7.7	70		38		12.6	..	96
7.8	73		41		12.7	..	97
7.9	75		44		12.8	..	97
8.0	77		47		12.9	..	97
8.1	79		50		13.0	..	97
8.2	81		52		13.1	..	97
8.3	82		54		13.2	..	97
8.4	83		56		13.1	..	97
8.5	84		57		13.4	..	97
8.6	85		59		13.5	..	98
8.7	86		60		13.6	..	98
8.8	87		62		13.7	..	98
8.9	88		64		13.8	..	98
9.0	89		66		13.9	..	98
9.1	90		68		14.0	..	98
9.2	91		70		14.1	..	99



$Sk = .98$ , based on data compiled from 652 male applicants used in computing the norms printed below. Considering that a low score (in minutes) denotes greater proficiency than a high score, the distributions are actually skewed negatively, which may be ascribed to the test being easy for the sampling of men with mechanical interests on whom it is used.

### *Summary*

Screwboard #1, a mechanical performance test, was devised to measure a portion of the abilities required to perform efficiently in certain mechanical jobs in a concern manufacturing business forms and systems. 'The test was determined to be reliable and valid if used as only one in a battery of tests. It was found to be easy for the group on which it was standardized but this is not considered of practical importance since critical scores are used (7.55 minutes) This score excludes approximately one-third of the male applicants.

•



A TECHNIQUE FOR OBTAINING SUMMARY DATA  
FROM AGGREGATE WEIGHTING SHEETS  
ON THE GRAPHIC ITEM COUNTER  
ATTACHMENT TO THE TEST  
SCORING MACHINE

CLIFFORD P. FROELICH  
United States Office of Education

AND

ROBERT J. KELLER  
University of Minnesota

ONE of the most time-consuming tasks in any statistical analysis is the transformation of raw data into some organized form. This task may be performed in a variety of ways but the end result is usually some form of summary data, which, in this article, includes sums, sums of squares, and sums of cross-products. The outstanding efficiency of punch-card techniques in obtaining such summary data has sometimes caused the research worker to evaluate the amount of labor required by his project in terms of direct calculator techniques or Hollerith punch cards. However, as a piece of equipment for tabulating data the Graphic Item Counter in terms of efficiency forms an intermediary link between the calculator and punch card equipment.

The usefulness of the test scoring machine in the tabulation of data has already received some attention. One of its earliest uses was described by Kuder (3) who outlined a method for obtaining correlation coefficients from coded scores on regular answer sheets. Bloom and Lubin (1) further added to the usefulness of the test scoring machine for this purpose by presenting methods in which the Graphic Item Counter attachment is used. Among other things, this article outlined a technique for securing summary data to be used in computing Pearson product-moment correlation coefficients. In both of these articles

geometrically coded scores were used and checks were developed in terms of the geometric progression. DuBois (2) used the counting sorter in a similar manner. McNamara and Weitzman (4) discussed the economy of the Graphic Item Counter for item analysis.

The purpose of this article is to outline a technique for obtaining summary data by the use of Aggregate Weighting Sheets, the Graphic Item Counter, and a specially prepared computational chart. This technique makes it possible to handle data without coding, checking becomes an automatic part of the total process and is introduced at successive intervals insuring accuracy at all stages, and the number of steps is decreased and the operations required are reduced to a comparatively low clerical level. In comparison with direct calculator computation this technique is superior in the amount of time saved and is equally efficient in accuracy provided four or more variables are used, the sample is fairly large (200 or more), and variables are expressed in one or two digits. The technique is unwieldy and time-consuming when the sample is less than 100 or when there are fewer than four variables expressed in one or two digits. In actual practice the number of variables which can be treated effectively is limited inversely by the magnitude of scores. Double-wiring of the machine and the use of commoning sheets partially offsets this limitation, and since variates can usually be expressed in two digits, this objection is not serious.

Because the Graphic Item Counter is well described in the literature and publications of the International Business Machines Corporation, no description will be included here.

The Aggregate Weighting Sheet is a modification of the standard answer sheet and contains 30 blocks, each of which has two horizontal rows of ten answer spaces. When two-digit numbers are entered on this sheet, the upper row is used to record the tens-digits of the variable, and the lower line, the units-digits. Variates are entered in cumulative form. For example, a score of 49 will be recorded by drawing a line through the answer spaces 1 through 4 in the upper row of a given block and a second line across spaces 1 through 9 in the lower row of

the same block. Since there are only 30 blocks on each sheet, the effective limit of this technique is 60 one-digit variables or 30 two-digit variables, each variable being assigned the same position for all individuals. With the use of a plastic stencil, two clerks can enter approximately 600 two-digit scores per hour. Each entry should be checked for accuracy and for quality of marking.

The Graphic Item Counter board is set up on the basis of entries on the Aggregate Weighting Sheet in such a way that the number of entries in each position can be tabulated. As an example, positions 1 through 9 on the tens line of variable I of the Aggregate Weighting Sheet are wired to counters 1 through 9 respectively. Each position in the units line of this variable is wired next to counters 10 through 18. Similarly, the remaining counters are wired to answer positions on the Aggregate Weighting Sheet in a predetermined and orderly manner.

Since much of the discussion which follows concerns the treatment of two-digit variables, it seems wise at this point to note the relationships inherent in their use. When a two-digit score is broken down into its integral parts, it is reconstructed as follows: (2) (Subscripts  $u$  and  $t$  refer to units and tens respectively throughout this article.)

$$X = 10X_t + X_u$$

When  $N$  individual two-digit scores are included, their sum is obtained by the relationship

$$\sum_{t=1}^{t=N} X = 10 \sum_{t=1}^{t=N} X_t + \sum_{t=1}^{t=N} X_u$$

and the other relationships required to obtain summary data for two-digit variables  $X$  and  $Y$  can be listed as follows:

( $\sum_{t=1}^{t=N}$  is denoted as  $\Sigma$ ).

$$\Sigma X^2 = 100 \Sigma X_t^2 + 20 \Sigma X_u X_t + \Sigma X_u^2$$

Note:

$$\Sigma Y = 10 \Sigma Y_t + \Sigma Y_u$$

$$\Sigma Y^2 = 100 \Sigma Y_t^2 + 20 \Sigma Y_u Y_t + \Sigma Y_u^2 \quad \text{and} \quad \Sigma X_t X_u = \Sigma X_u X_t$$

$$\Sigma XY = 100 \Sigma X_t Y_t + 10 \Sigma X_t Y_u + 10 \Sigma X_u Y_t + \Sigma X_u Y_u$$

After variables for each individual have been entered on a separate Aggregate Weighting Sheet and checked, the entire sample is sorted into sub-groups corresponding to entries on the

first line. When one-digit variables are used, the first sort will be on the first variable and when two-digit numbers are employed, the first sort will be made on entries in the tens row of variable I. It is important at this point to recognize that a zero sort will be made if one is required. For example, if the initial sort is made on the basis of entries in the tens line only, a value of 9 would be sorted into the zero-tens sub-group; scores from 10 through 19 will be sorted into the one-tens sub-group; and a score of 29 would appear with other scores ranging from 20 to 29 in the two-tens sub-group.

A reading taken on a Graphic Item Counter graph is taken at the end of each sub-group run, i.e., a new sheet is used to record G.I.C. marks for each sub-group. Each graph must be appropriately labeled to indicate the variable, the sort, and the sub-grouping on the sort. The notation, I-tens-3, would indicate variable I, tens-sort, sub-group 3. Such a notation provides sufficient information to identify all entries on the graph and is used throughout this discussion.

After all sub-groups of a given sort have been run through the machine, the entire sample is re-sorted on the basis of the next line of entries on the Aggregate Weighting Sheet. The process of running separate sub-groups through the machine and taking their respective readings is repeated until all of the variables have been sorted and run.

Several checks may be applied during the process to test the accuracy of sorting and of machine summation. First, the frequency of the count indicated in the total columns of each graph will agree with the number of Aggregate Weighting Sheets in the specific sub-group corresponding to that graph. Secondly, the counters attached to the sub-group sort will record exactly the number of counts in that sub-group. For example, when I-tens-0 is run through the machine, there will be no counts tallied on any counters attached to the tens line of variable I on the Aggregate Weighting Sheet. Similarly, I-tens-1 will appear as a single column of frequencies equal to the total frequency for the counter attached to the one position. Additional sub-groups will be represented by columns as high as the total column and equal in number to the sub-group it represents.

This circumstance usually makes it possible to identify the variable, sort the sub-groups directly upon study of the graph and knowledge of counter wiring. Thirdly, the cumulative recording of variables yields an additional check in that no bar is higher than that of the total and that the record of each set of counters attached to a given line on the Aggregate Weighting Sheet will appear as bars of equal or decreasing length. Exceptions to either one or all of these checks will indicate that one or more errors have been made. The skilled machine operator can sort and run approximately 300 sheets per hour and will soon be able to spot errors and isolate their cause with a minimum number of re-runs and/or re-sorts.

The graphs provided by the Graphic Item Counter record in bar form the frequency with which any given position on the Aggregate Weighting Sheet is marked. These values must be tallied in numerical form for computational purposes. The computational chart illustrated in Figure I is designed for this purpose. For convenience the variates used in this computational chart are referred to as "sort" and "count" variables. These terms can be defined as follows: "Sort" refers to that variable, or such part of it, e.g., units of a two-digit score, upon which sub-groupings are determined. "Count" denotes that variable, or such part of it, which is tallied by the counters of the G.I.C. attachment. The general notations, "A" and "B," have been employed on the computational charts to differentiate the "sort" from the "count" variables respectively.

The process of transcribing frequencies from graphs to computational charts is facilitated when both graphs and charts are arranged in a similar and orderly manner. Graphs of a given sort should be fastened together in sequence, i.e., sub-groups 0, 1, 2, 3, . . . 9, for each tens and units sort and combined by variables in the same progressive manner in which they are recorded by counters on the graphs themselves. In general, the order of the computational charts follow a similar pattern with this exception: each sort variable will be accompanied by a count variable, and only one sort and one count variable will be recorded on a single computational chart. The number of computational charts required by a given study will depend on

the size and number of variables. Cross-products require as many charts as there are digits raised to its own power ( $d^d$ ). Thus complete summary data for six two-digit variables would require 144 computational charts, half of which would be counterparts of other charts and used only for checking purposes.

Once the graphs and computational charts have been put into an ordered series as suggested above, entries on the charts can be made directly from the graphs. Sub-groups appear in column A at the left of the chart and range from 0 through 9. When computational charts and sections of the graphs have been matched, the counter frequencies from the graph will be recorded horizontally across the chart in cells 1 through 9 in the row designated by the sub-group. If entries on a single graph are read in order, these entries will be recorded on the same row of a series of charts corresponding to the appropriate counter variables. Sub-group 1, for example, will be recorded across row 1 of each chart in cells 1 to 9. Note that, although no counter can be attached to zeros, the entries in the zero column of the chart will be the same as the total bar on the graph. This is true because the total bar is the only measure of frequency which includes zeros and since scores are cumulative, the total frequency is the appropriate entry for the zero column. If desired, the actual frequency of zeros can be determined by finding the difference between the 0-column and the 1-column of the chart. This column actually represents the frequency distribution for the particular sort. When single digit variables are employed, this column is the frequency distribution of that variable. When two or more digit variables are used this column provides frequencies on but one of the digits and the total frequency distribution cannot be recovered from these values alone.

The cumulative nature of the counter record is utilized in the solution of the computational charts. When the sort variate is also expressed in cumulative form, the charts can be solved for all inherent summary data by simple addition. This speeds up the solution of computational charts by using addition throughout and makes adding machine or Comptometer tech-



niques applicable. The generalized principles used in the solution of charts can be expressed as follows.

The sum of a given set of values above their lowest value (or an assumed mean) is equal to the sum of their cumulated frequencies expressed as deviations above this value. When the lowest value is zero, the sum of the cumulated frequencies above zero is the total sum expressed in raw deviate form above zero. This can be expressed mathematically as follows:

$$\sum_{i=1}^i X = \sum_{i=1}^{i=h_x} CfX$$

where  $i$  refers to an individual;  $N$  is the size of the sample;  $X$ , the variate; and  $\Sigma$  indicates summation;  $CfX$  in the situation above represents the cumulated frequencies of the variate  $X$  from its highest value,  $h$ .

The sum of a set of cross-products is equal to the sum of their cumulated frequencies on one variable, cumulated and added with respect to deviations on the second variable, or, in terms of previous notation,

$$\sum_{i=1}^i XY = \sum_{i=1}^{i=h_x} C \cdot \sum_{i=1}^{i=h_y} CfY = \sum_{i=1}^{i=h_x} C \cdot \sum_{i=1}^{i=h_y} CfX$$

Whenever two variables are the same, as they are in this technique when they are both sorted and counted on the same variable, the sum of the cross-products is equivalent to the sum of squares, or, when  $X_i = Y_i$ ,

$$\sum_{i=1}^{i=N} XY = \sum_{i=1}^{i=N} X^2 = \sum_{i=1}^{i=N} Y^2$$

When both variates are expressed in continuous form or as single digit numbers on computational charts, summary data are obtained directly from the above relationships. When variates are broken down into tens and units, summary data are transformed back into linear form and obtained by algebraic relationships previously noted.

Before summary data for units and tens variables can be combined to obtain summary data for the total sample, it is necessary to compute these values from individual computational charts. The steps in this process are listed below in detail and, as a further aid, the computational chart shown in Figure I has been completely solved.



through 1 is  $\Sigma A$ . If the sort is a units sort, this  $\Sigma A$  is  $\Sigma X_u$ .  $\Sigma A$  is checked by cumulating the products of each entry in the zero column and its corresponding deviation in column A, or, as indicated on the chart, by obtaining  $\Sigma A f A$ . In Figure I,  $\Sigma A$  is  $\Sigma X_i$  and  $N$  is 35.

In like manner the counter variable is added in row  $\Sigma f B$  except that in this case, because  $B$  is a counter variable, the counters have already listed frequencies in cumulative form. Consequently it is only necessary to add each column and to record individual sums in row  $\Sigma f B$ . Entries in this row are added to obtain  $\Sigma B$ . The latter value is checked by comparing it with the final entry in column  $C \Sigma B$ , the purpose of which is explained in the following paragraph. In Figure I,  $\Sigma B$  is  $\Sigma Y_u$ .

The sum of the cross-products,  $\Sigma AB$ , is secured by adding the frequencies (which are already in cumulative form) in each row for every column except the zero ( $f A$ ) column, and entering the sums in column  $C \Sigma B$  at the extreme left of the chart. These sums are themselves cumulated and the final value listed in the zero cell of this column is  $\Sigma B$ . This must be true for in effect only deviations on the count variable have been added.  $\Sigma AB$  is equal to the sum of the cumulative frequencies in this column,  $C \Sigma B$ , down to but excluding the zero or  $\Sigma B$  cell. The cumulative nature of the counted frequencies is again used to obtain the cross-products in the opposite direction. These values are entered in row  $\Sigma d A$  at the bottom of the chart. In this instance, however, it is necessary to multiply each cell frequency in a given column by its respective sort deviate listed in column A. When the sums in this row,  $\Sigma d A$ , are added, the result will check with the previously obtained  $\Sigma AB$ . This value in Figure I is  $\Sigma X_i Y_u$ .

A check has thus been provided for each obtained value on the computational charts. Additional checks will now be introduced, to test the over-all accuracy of the entire process.

Since each variable is both sorted and run through the machine, there is available both a sort and a count record on each variable or part of a variable. Consequently the sum of variables cumulated by the machine as counter variable  $B$  should always agree with cumulated sort frequencies on the

same variable whenever it is listed as sort variable  $A$ . In practice, this means that all entries in column  $CfA$  will check with entries in row  $\Sigma fB$  in their entirety if  $A$  or  $B$  represent the same variable on a single chart or on any pair of charts. Similarly, under this condition, cumulative cross-products or like variables will check with each other, i.e., each entry in column  $C\Sigma B$  will agree with a comparable entry in row  $GdA$ .

Since each computational chart has a counterpart, inasmuch as each variable appears both as a sort and a count variable, it is unnecessary to solve completely all of the charts. The second chart serves solely as a check. As a result, to check the accuracy of the Graphic Item Counter count and its transcription to the chart, it is necessary only to solve the  $CfA$  column and the  $\Sigma fB$  row on both charts. This check will detect all but compensating errors. The latter would require a complete solution of both charts.

If columns and rows do not check as outlines above, it is still possible to isolate and define an error in count or transcription. In this case the matching is corrected in terms of the  $CfA$  column rather than the  $\Sigma fB$  row because of the twofold nature of the check on the sort variable outlined previously. The exact magnitude and cell location of the error can further be isolated by subtracting out cumulative entries to obtain actual frequencies for the cells. If a high degree of accuracy is maintained in all steps of the process, the vast majority of charts will check immediately. A few may contain one or two isolated errors which can be detected and corrected by this latter check.

### *Summary*

It has been the purpose of this article to outline a technique for obtaining summary data on a group of variables using raw scores entered on Aggregate Weighting Sheets and the Graphic Item Counter attachment for International Test Scoring Machines. The final way in which the summary data are gathered will depend of course upon the type of statistical analysis required by the problem. The method which has been outlined can be adapted to all techniques requiring full or partial summary data as previously defined to include sums, sums of squares, and cross-products.

The advantages inherent in the use of this technique appear to be the following:

1. The technique is superior to direct calculator techniques for samples of 200 or more cases. It is equally efficient for samples as small as 100 cases.

2. Variables are entered in raw score form and coding is unnecessary.

3. A computational chart has been developed for use with the Graphic Item Counter to obtain summary data.

4. The method contains complete checks of varied nature and introduced at sufficient places in the process to insure the accuracy of the Graphic Item Counter count and solutions for computational charts.

5. Steps in the application of the technique are sufficiently simple that they can be taught to and completed by clerks.

#### REFERENCES

1. Bloom, B. S. and Lubin, A. "Use of Test-Scoring Machine and the Graphic Item Counter for Statistical Work." *Psychometrika*, VII (1942), 233-241.
2. DuBois, P. H. "Some Statistical Operations on the Counting Sorter." *Psychometrika*, VI (1941), 383-390.
3. Kuder, G. F. "Use of the International Test Scoring Machine for the Rapid Computation of Tables of Intercorrelations." *Journal of Applied Psychology*, XXII (1938), 587-596.
4. McNamara, W. J. and Weitzman, E. "The Economy of Item Analysis with the IBM Graphic Item Counter." *Journal of Applied Psychology*, XXX (1946), 84-90.



# TABLE FOR USE IN THE COMPUTATION OF STATISTICS OF DICHOTOMOUS AND TRUNCATED DISTRIBUTIONS

E. K. TAYLOR

R. H. GAYLORD

Personnel Research Section, A.G.O.<sup>1</sup>

RECOGNITION of the fact that certain values recur frequently in computational statistics has led from time to time to the publication of a variety of tables designed to facilitate the computation of many common statistics. Tables of ordinates and areas of the normal curve are now to be found in almost every text book on statistics. Other tables dealing with proportion statistics have made their appearance in the literature from time to time. It is the aim of the present paper to bring together into a single table a number of often used values including some not readily found in statistical texts. This table is designed to aid the experimenter in the computation of such statistics as certain of the biserial correlation coefficients, the standard error of the proportion and the standard error of the difference between proportions. While it is undoubtedly true that certain workers will find the values applicable in instances other than those named, it is mainly for these statistics that the table is presented.

## *I. Organization of the Table*

The argument of the table is in terms of  $p$ , the proportion of the area of a unit normal<sup>2</sup> curve from a point of cut on the abscissa  $t$  to plus infinity. Thus for a distribution from which the lowest 10% has been eliminated, the value of  $p$  is equal to

---

<sup>1</sup> The opinions expressed in this paper are those of the authors and are not to be interpreted as representing official War Department policy.

<sup>2</sup> A unit normal curve is one where the total area is unity, the mean is zero and the standard deviation, unity.

.90 while one in which the concern is with the upper 15% the value of  $p$  would be .15. When dealing with a segment of a distribution truncated at both extremes,  $p_1$  will refer to the lower limit of the retained segment and will equal one minus the lower tail.  $p_2$  under such circumstances will refer to the upper limit of the retained portion and will be equal to the per cent of cases in the area above that limit.

As a result of the symmetry of the normal curve about its mid-point ( $p = .50$ ), a number of the entries in the table have identical values for the two values of  $p$ . The columns in this case have been given alphabetical designations. The values of  $p$  are printed in bold face. These values occur in two columns. In the left hand column labeled  $p'$  are values of  $p$  from .01 to .50 in ascending order. In the right hand column labeled  $p'$  are values of  $p$  from .50 to .99. The entries in the columns between these two bold-face columns are common to both of them.

The columns in which the entries have a unique value for each value of  $p$  have been split in two. The portion of such a column that is concerned with values of  $p$  from .01 to .50 is printed to the left of the bold-faced column of  $p$ . The portion of each such column associated with values of  $p$  from .50 to .99 is printed to the right of the second column of  $p'$  values. The columns have two-digit numerical designations. Those columns identified by numbers whose first digit is zero refer to the left  $p$  column, while those whose first digit is one refer to the right  $p'$  column. The second digit identifies the column. Thus the designation of the columns in which the means of the tail of a distribution are recorded are 01 and 11. Column 01 gives the mean of the tail of distributions in which the tail contains 50% or less of the cases and column 11 the mean of tails containing 50% or more of the distributions.

## II. Columns 01 and 11—The Mean Value of the Tail of the Distribution— $\left(\frac{y}{p}\right)$

Entries in this column give the mean values in terms of standard scores of a tail of a normal distribution. Suppose for example it were necessary to know the mean score of the high-



est 25% of students on a history test when the mean of the entire population was known to be 85 and the standard deviation 10. The value of  $y/p$  for  $p = .25$  is found to be equal to 1.2711. This value is a standard score. The relationship between a standard score and a raw score may be expressed as

$$(1) \quad \sigma t + M = X$$

where  $X$  = a raw score

$M$  = the mean

$\sigma$  = the standard deviation and

$t$  = a standard score

In the example above  $10(1.2711) + 85 = 97.711$  which is the raw score mean of the highest 25% of the population. If we were interested in the mean of the *lower* 75% of this same distribution we would obtain our  $y/p$  value from column 11 opposite the  $p$  entry of .75. Since, however, the mean required is of a *lower* tail of the distribution where more of the standard scale values are negative,  $y/p$  must be given a negative sign. Applying formula (1) would yield  $10(-.4237) + 85 = 80.763$  which is the mean of the lowest 75% of the population. Note that the  $p$  value to be used as the argument here is the proportion of cases in the tail for which the mean is being computed and that the sign of the mean standard scale score ( $y/p$ ) is positive when the tail is at the upper end of the distribution and negative if at the lower end.

Another situation in which it is useful to be able to determine the relationship between the mean of a tail of a normal distribution and the entire distribution is where data above or below a certain point are not available. We might take as an example an instance in which 18% of a population made a perfect score on a nation-wide arithmetic test. If we wished to compare this group with another which did not fare so well, it would be illogical to assume that all of the 18% whose ability was beyond the range of the test had only as much arithmetic competence as the test was able to measure. A far better approach would be to estimate (from the mean of the 82% of the cases that fell on the scale of the test) what the mean of the total group would be if the continuum were extended throughout the range of ability. Since there are 82% of the cases in the

tail of the distribution for which data are available, we find from column 11 that the mean of this tail deviates by .32 standard deviations from the mean of the total distribution. Since it is a lower tail with which we have been concerned, the sign of the deviation is of course negative. We now have our answer but since it is in terms of the standard deviation of the total distribution, we must determine that statistic before that answer is of any value to us. It is to this task that we now turn.

### *III. Columns 02 and 12—The Standard Deviation of the Tail of a Normal Curve*

By mathematical manipulations beyond the scope of this paper, the ratio of the standard deviation of the tail of a normal curve to that of the entire curve may be determined by the equation

$$(2) \quad SD_t = \sqrt{1 + \frac{(ty)^2}{p} - M_t^2}$$

Values of equation (2) for each value of  $p$  are recorded in columns 02 and 12. In the problem presented in the previous section, there were 82% of the cases in the tail of the distribution for which data were distributed. Looking in column 12 in the row of  $p = 82$ , we find that the ratio of the standard deviation of this tail to that of the complete distribution is .78. Suppose that by actual computation we found the mean and standard deviation of the 82% tail to be 56.84 and 5.28 respectively. The value 5.28 would represent 78% of the standard deviation of the total distribution. This is found to be 6.77. Knowing further that the mean of the total distribution is .32 standard deviations above that of the lower tail, we now find this difference to be 2.17 score points. Adding this amount to the mean (56.84) of the 82% of the cases that did not make a perfect score, we obtain 59.01 as the mean of the total distribution.

### *IV. Columns 03 and 13—Aid to the Computation of $r_{bis} - \left(\frac{p}{y}\right)$*

The biserial correlation coefficient yields the relationship between two normally distributed variables one of which has

been dichotomized so that there are  $p$  per cent of the cases in the higher group. One of the formulas by means of which the biserial correlation coefficient may be determined is

$$(3) \quad r_{bis} = \frac{M_h - M_t}{\sigma_t} \cdot \frac{p}{y} \quad \text{where}$$

$r_{bis}$  = biserial correlation coefficient

$M_h$  = the mean value of the continuous variable for those cases in the high dichotomy

$M_t$  = the mean of the total group on the continuous variable

$\sigma_t$  = the standard deviation of the total group on the continuous variable

Suppose the history test referred to above had been administered at the beginning of a school course and it is desirable to determine if prior knowledge of history is correlated with the ability to pass the course. If 59% of the students passed the course and if the mean test score of these was 90, the correlation between the test and success in the course would be according to equation (3),

$$r_{bis} = \frac{90 - 85}{10} (1.518) = .76$$

The coefficient is determined by computing the mean and standard deviation of the continuous variable for the total group and the mean of the continuous variable for those cases in the high group. The difference between the two means is divided by the standard deviation. This quotient is multiplied by the proper value of  $p/y$ , (where  $p$  is the proportion of cases with the higher score in the dichotomous variable) to secure the biserial. The sign of the correlation is determined by the sign of the difference between the mean score of the high and total groups.

#### V. Columns 04 and 14—Aid to the Compu-

$$\text{tation of } r_{pt\ bis} - \left( \sqrt{\frac{p}{q}} \right)$$

Richardson and Stalnaker (4) were first to point out that no justification exists for the assumption that certain variables are normally distributed. In other cases it was pointed out that the continuous nature of the variable could not be defended. Under such circumstances, the computation of the

biserial correlation is of course obviously inappropriate.<sup>3</sup> As an alternative, Richardson and Stalnaker have proposed the *point-biserial correlation*. This statistic is a Pearson product-moment correlation in which there are but two class intervals in one of the variables. The point-biserial correlation may be used in any situation in which numerical values may be assigned to the two categories of the dichotomous variable. For convenience, values of zero and unity are generally assigned to the low and high groups respectively. The assignment of *any* other values to the categories would have no effect on the point-biserial correlation coefficient.

Thus it may be seen that if the arbitrary values  $X$  and  $Y$  are assigned to the two categories, they may be treated as 1 and 0 by subtracting  $Y$  from each and then dividing by  $(X - Y)$ . The point-biserial being a product-moment correlation is unaffected by this linear transformation.

The point-biserial correlation coefficient may be computed by the formula

$$(4) \quad r_{pbis} = \frac{M_h - M_l}{SD_t} \cdot \sqrt{\frac{p}{q}}$$

where the symbols are the same as in formula (2).

The computation of the point-biserial correlation coefficient exactly parallels that of the biserial except for the last step. In the case of the point-biserial the quotient  $\left(\frac{M_h - M_l}{SD_t}\right)$  is multiplied by  $\sqrt{\frac{p}{q}}$  (where  $p$  is again the proportion of cases in the higher category on the dichotomized variable).

#### VI. Column A. Aid to Alternate Method of

$$\text{Computing } r_{bis} = \left(\frac{pq}{y}\right)$$

An alternate formula for the computation of the biserial correlation coefficient is

$$(5) \quad r_{bis} = \frac{M_h - M_l}{SD_t} \cdot \frac{pq}{y} \quad \text{where}$$

<sup>3</sup> The writers recognize the inadequacy of these justifications for the use of the point-biserial. The advanced reader is referred to a complete discussion of the biserial coefficients in a paper soon to be published by H. E. Brogden.

- $M_h$  = the continuous variable mean of the members of the sample in the high group on the dichotomous variable  
 $M_l$  = the continuous variable mean of the members of the sample in the low group on the dichotomous variable  
 $SD_t$  = Standard deviation of the continuous variable for the entire population  
 $p$  = proportion of cases in the high group  
 $q$  = proportion of cases in the low group  
 $y$  = the ordinate of the normal curve at the point  $p$

In using equation (5) the ratio of the difference between the mean of the high and low group to the standard deviation of the total group would be computed. This ratio would be multiplied by the appropriate entry in column A to yield the biserial correlation coefficient. The sign of the coefficient would, of course, be the same as that of the difference between the two means, i.e., positive if the mean of the high group is larger and negative if the mean of the low group is larger.

#### VII. Column B—Values of $\sqrt{pq}$

The value of  $\sqrt{pq}$  has several uses in the treatment of statistical data. The reader in working with percentage statistics will have occasion to refer to it frequently.

*A. Computation of the Point-Biserial.*—As in the case of the biserial, the point-biserial correlation may be computed from the difference between the continuous variable means of the high and low groups. The equation is

$$(6) \quad r_{pbis} = \frac{M_h - M_l}{SD_t} \cdot \sqrt{pq}$$

where the symbols have the same meaning as in equation (5). Computation parallels that of the biserial.

*B. The Standard Error of a Two-Point Distribution.*—This is given by the formula  $\frac{\sqrt{pq}}{\sqrt{N}}$  when the distribution is of unit

area. When  $p$  is the proportion of cases falling into any given score category on any variable and there are  $N$  cases in the total sample, the number of cases in that category is equal to  $Np$  and the standard error of that frequency is equal to  $\sqrt{Npq}$ .

If for example 50% of students in a sample of 100 make scores between 50 and 60 on an achievement test, the standard

error of that score range is equal to  $\sqrt{100} \sqrt{pq}$  or  $10(.25)^{.5}$  or 5. This may be interpreted to mean that in successive samples of the same size the probability is less than one in one thousand that there will be fewer than 35 or more than 65 cases in this step interval.

*C. Standard Error of a Percentage.*—It follows directly from the preceding section that the standard error of a percentage is equal to  $100 \frac{\sqrt{pq}}{\sqrt{N}}$ . This statistic is interpreted in the same manner as the standard error of a frequency. Note should be made, however, of the fact that the distributions of proportions, percentages and frequencies are markedly skewed in the vicinities of the extremes, i.e., for proportions zero and unity. This is particularly true when  $N$  is small. As a general rule the usual standard error interpretations are considered inapplicable when  $p$  is less than .05 or greater than .95, or where  $N$  is less than 100. The reader is referred to Lindquist's "A First Course in Statistics" (1, pp. 125-129) for a clear exposition of the methods of testing the significance of proportion statistics.

### VIII. Column C. Values of $pq$

The values of  $pq$  are included in these tables because they so frequently appear as parts of equations whose form is such that it is not feasible to present them in complete tabular form. Most of these statistics are concerned with the difference between proportions.

The equation for the difference between two proportions from uncorrelated samples is

$$(7) \quad SD_{p_1 - p_2} = \sqrt{\frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}} \quad \text{where}$$

$p_1$  = the proportion of cases in sample one having the characteristic under consideration

$q_1 = 1 - p_1$

$N_1$  = number of cases in sample 1

$p_2$  = the proportion of cases in sample two having the characteristic under consideration

$q_2 = 1 - p_2$

$N_2$  = the number of cases in sample 2

When there are an equal number of cases in the two samples, (7) reduces to

$$(8) \quad SD_{p_1-p_2} = \sqrt{\frac{p_1q_1 + p_2q_2}{N}} \quad \text{where}$$

$N$  = the number of cases in each sample

For example, suppose that vocational guidance was given to 100 male students and 100 female students. Fifty-six of the female students and 62 of the male students entered the vocation recommended by the counselor. It is desired to know if the difference obtained was representative, or could be explained as a result of sampling fluctuations. The difference between proportions would be

$$.62 - .56 = .06$$

and the standard error of this difference according to equation (8) would be

$$(9) \quad \sqrt{\frac{.2356 + .2464}{100}} = \sqrt{\frac{.4820}{100}} = \sqrt{.0048} = .069.$$

Since the ratio of the difference to its standard error is somewhat less than one, the hypothesis that the difference could be explained by chance is not refuted by the data.

#### IX. Column D. Conversion of $r_{pt\ bis}$ to $r_{bis}$ and

$$\text{Standard Error of } r_{bis} = \left( \frac{\sqrt{pq}}{y} \right)$$

The ratio of the point-biserial to the biserial correlation coefficient is a function of the proportions into which the sample is split on the dichotomous variable. The point-biserial is always smaller than the biserial. Point-biserial correlations may be converted to biserials by the formula

$$(10) \quad r_{bis} = r_{pt\ bis} \cdot \frac{\sqrt{pq}}{y}$$

Since the factor  $\frac{\sqrt{pq}}{y}$  is a maximum for  $p = .50$  and decreases as  $p$  approaches either zero or unity, it is apparent that the ratio of the point-biserial to the biserial is largest when the dichotomous variable contains 50% of the cases in each category.

Another statistic employing the values in this column is the

standard error of a biserial correlation. Peters and Van Voorhis (3) give the following equation

$$(11) \quad SD_r = \frac{\frac{\sqrt{pq}}{y} - r^2}{\sqrt{N}} \quad \text{where}$$

$SD_r$  = standard error of a biserial correlation

$\frac{\sqrt{pq}}{y}$  = the entry in column D appropriate to the split on the dichotomous variable

$r^2$  = the square of the obtained value of the biserial correlation

$N$  = the number of cases on which the correlation is based

In testing whether or not an obtained biserial correlation differs significantly from zero (11) reduces to

$$(11a) \quad SD_r = \frac{\frac{\sqrt{pq}}{y}}{N}$$

Formulas (11) and (11a) are of limited usefulness, particularly when  $p$  is less than .10 or greater than .90. Formula (11) is further limited for values of  $r_{bu}$  greater than .5. The reader should not attempt to apply these statistics without first consulting some good statistics book on the general topic of the reliability of correlations. Lindquist's (2) "Statistical Analysis in Educational Research" is recommended.

#### X. Column E. Conversion of $r_{bu}$ to $r_{pt}$ via $\left(\frac{y}{\sqrt{pq}}\right)$

Occasionally studies appear in the literature reporting biserial correlation coefficients when the assumptions on which this statistic is based have been violated. Under these circumstances, the reader may desire to convert the reported coefficients to point-biserials. If the value of  $p$  is given, this may be accomplished by the equation

$$(12) \quad r_{pt \text{ via }} = r_{bu} \cdot \left(\frac{y}{\sqrt{pq}}\right)$$

Thus if it is reported that 100 male and 100 female students were given a test of general intelligence and that the biserial correlation was found to be .40, a better measure of the relation-



ship may be determined by multiplying the obtained correlation by .7977. The point-biserial of .32 is free of the assumption that sex is a normally distributed variable and that a linear relationship exists between sex and intelligence throughout the complete range of each.

*XI. Column F. Ordinates of the Normal Curve—(y)*

The value of the ordinate of the unit normal curve ( $y$ ) may be found in almost any statistics text. It is presented here primarily for the convenience of having it available for use along with these other values in a single table and secondly because most statistics texts present it only with standard score units as the argument. This makes interpolation (a tiresome and somewhat inaccurate task) necessary if the ordinate corresponding to a given percentage of area is required.

These values will be most useful in the determination of the mean (in standard score terms) of a segment of the normal curve. In a normal curve cut at the points  $p_1$  and  $p_2$ , the standard score mean of the cases between the two points is computed by the formula

$$(13) \quad M_s = \frac{y_1 - y_2}{p} \quad \text{where}$$

$M_s$  = the mean (in terms of standard scores) of the segment of a unit normal curve

$p$  = the proportion of the cases in the segment under consideration

$y_1$  = the ordinate of the normal curve at the *lower* limit of the segment

$y_2$  = the ordinate of the normal curve at the *upper* limit of the segment

Suppose then an experiment was being conducted in which it was necessary to know the mean and standard deviation, but that the available data, while giving the distribution by year of those cases which had completed between 7 and 12 years of schooling, only indicated that 11% of the population had not completed the sixth grade while 13% had attended school beyond the 12th grade. The mean number of years completed for those subjects who had completed the seventh grade but had not gone beyond high school is found to be 10.13 years and the

standard deviation of this group is 1.44 years. What is the mean and standard deviation of the total population?

The first step in the determination of the answer is to find the relationship between the mean of the segment and that of the total population. This may be accomplished by employment of equation (13)

$$M_s = \frac{.1880 - .2115}{.76} = \frac{-.0235}{.76} = -.03$$

This of course answers our question only in terms of the deviation of the mean of the segment from that of the total population, i.e., the mean of the total population is .03 standard deviation points higher than that of the segment. Since the standard deviation under consideration is that of the entire population rather than of the segment, that standard deviation must be determined before the question raised can be completely answered. The procedure for this determination is given in the next section.

### *XII. Column G. Standard Deviation of a Segment of the Normal Curve—(ty)*

The ratio of the standard deviation of a segment of the normal curve to that of the complete normal curve is given by the formula

$$(14) \quad SD_s = \sqrt{1 + \frac{(ty)_1 - (ty)_2}{p}} - M_s^2 \quad \text{where}$$

$SD_s$  = ratio of the standard deviation of the segment to that of the complete distribution

$(ty)_1$  = the signed value of the entry in column G corresponding to the  $p$  value of the *lower* limit of the segment

$(ty)_2$  = the signed value of the entry in column G corresponding to the value of  $p$  at the *upper* limit of the segment

$p$  = the proportion of the population in the segment

$M_s^2$  = the square of the standard scale value of the mean of the segment

It should be noted that the sign of a value of a given entry in the  $ty$  column is determined by its  $p$ . In our problem of the education distribution  $(ty)_1$  is determined by the  $p$  value of the lower limit of the segment which is .11. The  $p$  value of the

upper limit of the segment which determines the sign of  $(ty)_2$  is .87. If the value of  $p$  is less than .50, the sign of  $(ty)$  is negative. If the value of  $p$  is greater than .50, the sign of  $(ty)$  is positive. Thus in our problem  $(ty)_1$  would be negative and  $(ty)_2$  positive. Substituting appropriate values in equation (14) we obtain

$$\begin{aligned} SD_s &= \sqrt{1 + \frac{.2206 - .2382}{.76} - (.03)^2} \\ &= \sqrt{1 - \frac{.4588}{.76} - .0009} \\ &= \sqrt{1 - .6036 - .0009} = \sqrt{1 - .6045} \\ &= \sqrt{.3955} = .6289 \end{aligned}$$

We now know that our obtained standard deviation of 1.44 years is 62.89% of the standard deviation of the total population. Applying this information we may obtain an estimate of the total standard deviation as follows:

$$\begin{aligned} SD_t(.6289) &= 1.44 \\ SD_t &= 1.44/.6289 \\ SD_t &= 2.29 \end{aligned}$$

Since it had previously been determined that the mean of the total population was .03 standard deviations above that of the segment, we may now compute the mean of the total population. Converting the standard score to a raw score difference by multiplying .03 by 2.29 we find that the mean of the total population is .0687 of a year higher than that of the segment. Adding this amount to the mean of the segment (10.13 years) and rounding to two places yields 10.20 years as the mean of the total population.

### *XIII. Column H. Standard Score Deviation at Point of Cut—(t)*

As in the case of the values of the ordinates of the normal curve associated with the areas indicated by  $p_1$ , the standard scale values may be found in all statistical texts and are reproduced as a matter of convenience. In using these values a negative sign is required if  $p$  is less than .50 and a positive sign if  $p$  is greater than .50. For example, if in a set of normally dis-

Table for Use in the Computation of Statistics of Dichotomous and Truncated Distributions\*

01	02	03	04	A	B	C	D	E	F	G	H	11	12	13	14
$\frac{y}{p}$	$\sigma r$	$\frac{p}{y}$	$\sqrt{\frac{p}{q}}$	$\frac{pq}{y}$	$\sqrt{pq}$	$pq$	$\frac{\sqrt{pq}}{y}$	$\frac{y}{\sqrt{pq}}$	$y$	$(ty)$	$t$	$\frac{y}{p}$	$\sigma r$	$\frac{p}{y}$	$\sqrt{\frac{p}{q}}$
26650	31	3752	1005	3714	0995	0099	37333	2679	0267	0621	23263	98	0269	3715	99499
24210	33	4131	1428	4048	1400	0196	28915	3458	0484	0994	20537	98	0494	2024	70000
22680	35	4409	1758	4277	1706	0291	25073	3988	0680	1279	18808	97	0701	1426	56862
21540	36	4642	2042	4456	1960	0384	22745	4397	0862	1509	17507	96	0898	1114	48990
20626	37	4848	2293	4606	2179	0475	21127	4733	1031	1696	16449	95	1086	9211	43589
19853	38	5037	2526	4735	2375	0564	19937	5016	1191	1852	15548	94	1267	7891	39581
19180	39	5214	2744	4848	2551	0651	19000	5263	1345	1982	14758	93	1444	6976	36450
18583	40	5381	2950	4951	2713	0736	18249	5480	1487	2089	14051	92	1616	6188	33912
18041	41	5542	3145	5043	2862	0819	17624	5674	1624	2177	13408	91	1784	5504	31798
17548	41	5698	3333	5128	3000	0900	17094	5850	1755	2249	12816	90	1950	5128	30000
17094	42	5850	3416	5206	3129	0979	16640	6010	1880	2306	12265	89	2113	4733	28445
16670	42	5999	3693	5279	3250	1056	16247	6155	2000	2350	11750	88	2273	4399	27080
16272	43	6145	3865	5346	3363	1131	15897	6290	2115	2382	11264	87	2431	4112	25869
15899	44	6290	4035	5409	3470	1204	15590	6414	2226	2405	10803	86	2588	3864	24785
15544	44	6433	4201	5468	3571	1275	15316	6529	2332	2417	10364	85	2743	3646	23805
15206	45	6576	4365	5524	3666	1344	15067	6637	2433	2420	9945	84	2896	3452	22913
14885	45	6718	4525	5576	3756	1411	14843	6737	2531	2415	9542	83	3049	3280	22016
14577	46	6860	4685	5625	3842	1476	14642	6830	2624	2402	9154	82	3200	3125	21344
14282	46	7002	4844	5671	3923	1539	14456	6917	2714	2383	8779	81	3350	2985	20647
13998	47	7144	5000	5715	4000	1600	14288	6999	2800	2356	8416	80	3500	2858	20000
13724	47	7287	5156	5756	4073	1659	14132	7076	2882	2324	8064	78	3648	2741	19396
13459	48	7430	5311	5796	4142	1716	13989	7149	2961	2286	7722	78	3796	2634	18829
13203	48	7575	5465	5832	4208	1771	13858	7216	3036	2243	7388	77	3944	2536	18297
12953	49	7720	5620	5867	4271	1824	13739	7279	3109	2196	7063	76	4090	2445	17795
12711	49	7867	5773	5900	4330	1875	13626	7339	3178	2144	6745	75	4237	2360	17320

01	02	03	04	A	B	C	D	E	F	G	H	11	12	13	14
$\frac{y}{p}$	$\sigma_r$	$\frac{p}{y}$	$\sqrt{\frac{p}{q}}$	$\frac{pq}{y}$	$\sqrt{\frac{pq}{y}}$	$pq$	$\sqrt{\frac{pq}{y}}$	$\frac{y}{\sqrt{pq}}$	$y$	$(y)$	$z$	$\frac{y}{p}$	$\sigma_r$	$\frac{p}{y}$	$\sqrt{\frac{p}{q}}$
12476	50	8016	5928	26	5932	4386	1974	7396	3244	2087	6433	74	4383	73	2281-
12446	50	8166	6082	27	5961	4440	1971	7447	3306	2026	6128	73	4529	72	2208-
12023	51	8318	6236	28	5989	4490	2016	7497	3366	1962	5828	72	4675	71	2139-
11803	51	8472	6391	29	6015	4538	2059	7543	3423	1894	5534	71	4821	71	2074-
11590	51	8628	6547	30	6040	4583	2100	7587	3477	1833	5244	70	4967	70	2013-
11380	52	8787	6703	31	6063	4625	2139	7628	3528	1750	4959	69	5113	70	1956-
11175	52	8948	6860	32	6085	4665	2176	7666	3576	1672	4677	68	5259	69	1902-
10974	53	9112	7018	33	6105	4702	2211	7702	3621	1593	4399	67	5405	69	1850-
10776	53	9279	7178	34	6124	4737	2244	7735	3664	1511	4125	66	5552	68	1801-
10583	54	9449	7338	35	6142	4770	2275	7765	3704	1427	3853	65	5698	68	1755-
10392	54	9623	7500	36	6158	4800	2304	7794	3741	1341	3585	64	5845	67	1711-
10204	55	9800	7664	37	6174	4828	2331	7820	3776	1253	3319	63	5993	66	1669-
10020	55	9980	7829	38	6188	4854	2356	7844	3808	1163	3055	62	6141	66	1628-
9838	55	1016-	7996	39	6200	4877	2379	7867	3837	1072	2793	61	6290	65	1590-
9659	56	1035-	8165	40	6212	4899	2400	7886	3863	978	2533	60	6439	65	1553-
9482	56	1055-	8336	41	6223	4918	2419	7905	3887	884	2275	59	6589	64	1518-
9307	57	1074-	8509	42	6234	4936	2436	7919	3909	789	2019	58	6739	64	1484-
9134	57	1095-	8686	43	6240	4951	2451	7933	3928	693	1764	57	6891	64	1451-
8964	58	1116-	8864	44	6247	4964	2464	7946	3944	596	1510	56	7043	63	1420-
8796	58	1137-	9045	45	6253	4975	2475	7956	3958	498	1257	55	7196	63	1390-
8629	58	1159-	9230	46	6258	4984	2484	7964	3969	398	1004	54	7351	62	1360-
8464	59	1181-	9417	47	6262	4991	2491	7971	3978	300	0753	53	7506	62	1332-
8301	59	1205-	9598	48	6264	4996	2496	7975	3984	200	0502	52	7662	61	1305-
8139	60	1229-	9802	49	6266	4999	2499	7978	3988	100	0251	51	7820	61	1279-
7979	60	1253-	10000	50	6267	5000	2500	7979	3989	000	0000	50	7979	60	1253-

\* ALL DECIMAL POINTS have been omitted from this table. Columns 02, 12, p and p' are 100 times their real value. All other columns are 10,000 times their true value. Dashes replace the units positions of those entries where the units position is not significant. These dashes should NOT be interpreted as minus signs.

tributed test scores having a mean of 100 and standard deviation of 10 was cut at  $p = .23$ , i.e., if the lowest 23% of the population was to be eliminated, the standard scale value of the critical score value would be  $-.7388$ . If on the other hand, 77% of the cases were to be eliminated, the standard scale value of the critical score would be  $+.7388$ . Such critical scale values may be converted to raw scores by the equation:

$$(15) \quad x = \sigma t + M_x \quad \text{where}$$

$x$  = the raw score of the variable

$\sigma$  = the standard deviation of the raw score distribution

$t$  = the signed standard score value

$M_x$  = the raw score mean of the distribution

#### REFERENCES

1. Lindquist, E. F. *A First Course in Statistics*. New York: Houghton-Mifflin Company, 1942, pp. 125-129.
2. Lindquist, E. F. *Statistical Analysis in Educational Research*. New York: Houghton-Mifflin Company, 1940, pp. 210-219, 243.
3. Peters, C. C. and Van Voorhis, W. R. *Statistical Procedures and Their Mathematical Bases*. New York: McGraw-Hill Book Company, 1940.
4. Richardson, M. W. and Stalnaker, J. M. "A Note on the Use of Biserial  $r$  in Test Research." *Journal of General Psychology*, VIII (1935), p. 461.

## MEASUREMENT ABSTRACTS

Ax, Alfred F. "A Validation Study of the Rotter-Jensen Level of Aspiration Test." *Journal of Personality*, XV (1946), 166-172.

On the basis of total case records evaluated as better integrated (A) or less integrated (B) by ten criteria independent of the *Level of Aspiration Test*, 25 subjects were divided into two clinical groups. It was found, on administration of the *Level of Aspiration Test*, that while Group A tended to scale estimates close to performance, Group B consistently either overestimated or underestimated achievement. The subgroup of Group B with large negative aspiration scores tended also to larger "Judgment Difference" scores, and the hypothesis is offered that the ability to predict a future performance accurately depends more on the subject's "sense of self-judgment" than on sheer extrapolation from past performance. *Frances Smith.*

---

Carter, Launor F. and Dudek, Frank J. "The Use of Psychological Techniques in Measuring and Critically Analyzing Navigators' Flight Performance." *Psychometrika*, XII (1947), 31-42.

Under controlled flight conditions, the distance between a navigator's report of position and his actual position is a criterion of success in dead reckoning navigation. Students' logs were evaluated for five separate missions by comparing the students' entries with standards determined by experts. The reliability of this technique is indicated by the fact that mission-to-mission intercorrelations of error scores were low, while the intercorrelations between legs of the same mission were moderately high. The intercorrelations between the error scores for the different navigation variables were computed and analyzed by using both factor analysis and multiple-regression techniques. Both analyses indicated that a major portion of all dead reckoning error could be attributed to errors made in determining magnetic deviation. As a result of these analyses, recommendations were made for changing the instruction in dead reckoning and for alterations in the equipment used. (Courtesy *Psychometrika*.)

---

Cattell, Raymond B. "Oblique, Second-order, and Cooperative Factors in Personality Analysis." *Journal of General Psychology*, XXXVI (1947), 3-22.

It is argued from the intrinsic correlational relationships of actual data that interpretation in factor analysis by means of oblique factors

offers the best possible solution of a given psychological problem. The relationship between any two such factors may, however, be slight, and their factor loadings in the same set of variables may yet be appreciably correlated. In such a situation the factors may be called cooperative factors, and the nature of their relationship demands further investigation of possible underlying psychological mechanisms. A second-order factorization of 12 primary personality factors previously reported upon is presented by alternative methods of analysis, and the nature of the factors found is discussed, with generalizations offered as to the nature of second-order factors. *Frances Smith.*

---

Cronbach, Lee J. "Test 'Reliability': Its Meaning and Determination." *Psychometrika*, XII (1947), 1-16.

The concept of test reliability is examined in terms of general, group and specific factors among the items, and the stability of scores in these factors, from trial to trial. Four essentially different definitions of reliability are distinguished which may be called the hypothetical self-correlation, the coefficient of equivalence, the coefficient of stability, and the coefficient of stability and equivalence. The possibility of estimating each of these coefficients is discussed. The coefficients are not interchangeable and have different values in corrections for attenuation, standard errors of measurement, and other practical applications. (Courtesy *Psychometrika*.)

---

Dubin, Elizabeth Ruch. "The Effect of Training on the Tempo of Development of Graphic Representation in Preschool Children." *Journal of Experimental Education*, XV (1946), 166-173.

The drawings of fifty-two pre-school children were collected and classified according to a scheme ranging from scribble to representation. An experimental group was then trained to attain the next higher level of drawing after which drawings were again collected and scored according to a system devised by Marion Monroe. The author concludes that training improved the development of skill in graphic representation and that an art program which could advance the drawing level of a child can be developed even for two-year-old children in nursery school. *Harold Morak.*

---

Engle, T. L. "The Use of The Harrower-Erickson Multiple-Choice (Rorschach) Test in Differentiating Between Well-Adjusted and Maladjusted High-School Pupils." *Journal of Educational Psychology*, XXXVII (1946), 550-556.

From a total of 6,558 students in several Indiana high schools, 111 boys and girls were selected by teachers as maladjusted, and 118 as outstandingly well-adjusted. The *Harrower-Erickson Multiple-Choice Rorschach Test* was administered to the two groups, the subjects being required to underline thirty responses, and then to under-



line as many more as they wished. The test was discriminating for the boys, but not for the girls. From the results of these two extreme groups, the author concludes that the test would not be of much value for detecting minor maladjustments in the great group of pupils near the center of the distribution. *Leroy S. Burwen.*

---

Goodman, C. H. "The MacQuarrie Test for Mechanical Ability: II. Factor Analysis." *Journal of Applied Psychology*, XXXI (1947), 150-154.

The author employs the Thurstone Centroid Method of factor analysis to determine the nature of the factors measured by the *MacQuarrie Test for Mechanical Ability*. He isolates three factors, two of which are clearly identified as a space factor and a controlled manual factor and the third tentatively identified as of a visual inspection nature. *Harold Mosak.*

---

Himmelweit, H. T., Desai, M., and Petrie, A. "An Experimental Investigation of Neuroticism." *Journal of Personality*, XV (1947), 173-196.

Eleven objective personality tests and a questionnaire were given to a group of 105 psychiatric service patients and 93 surgical service patients from the Southern Hospital, Dartford, England. The aim of the study was to provide a basis for building a battery of reliably discriminative tests of neuroticism and also to isolate by factorial analysis a general factor which might serve as the internal criterion of validity for the tests. The tests were found to discriminate significantly between normal and neurotic groups (C.R. 8). By means of factorial analysis of test results, details of procedure for which are unspecified, two factors were isolated, the first accounting for 18 per cent of the score variance, the second for 7 per cent. Correlations of test scores with internal and external validity criteria indicate a validity of about .80 for the battery. *Frances Smith.*

---

Irwin, F. W. and Gebhard, M. E. "Studies in Object-Preferences: The Effect of Ownership and Other Social Influences." *American Journal of Psychology*, LIX (1946), 633-651.

A series of experiments was carried out on groups of orphan children from the ages of five to nineteen to study the effect on object-preferences of such influences as ownership, preference of another person, having the object in hand as opposed to being told about it, etc. A majority of the children indicated a preference for their own objects over those of other children, whether their own objects were absent or present, and a preference for the objects of friends over those of other children. Praise by the experimenter enhanced the desirability of an object. Some of the other results were not conclusive. *Leroy S. Burwen.*

Kay, Lillian Wald. "Frame of Reference in 'Pro' and 'Anti' Evaluation of Test Items." *Journal of Social Psychology*, XXV (1947), 63-68.

This study tests the assumption that the acceptance of each statement indicated a negative affect towards Jews in the Levinson and Sanford scale for the measurement of anti-Semitism. Ten statements, six of "opinion" and four of "attitude" were chosen and administered to three selected groups. The results indicated that the acceptance of each statement did not always indicate a negative affect, but might be dependent on the frame of reference of the individual. *Leroy S. Burwen.*

---

Leitch, Mary and Schafer, Sarah. "A Study of the Thematic Apperception Tests of Psychotic Children." *American Journal of Orthopsychiatry*, XVII (1947), 337-342.

On the assumption that in the diagnosis of childhood psychosis corroborative data obtained from diagnostic psychological tests are especially useful, a study was made at the Menninger Clinic of the *Thematic Apperception Test* stories of 15 psychotic and 15 maladjusted non-psychotic children. The children were matched for age, with a range of from 5 to 17 years. The stories of the psychotic group were found to be differentiated from those of the non-psychotic group by the presence of severe disturbances in thought organization, by gross perceptual distortions, and by frequent references to violence or death. *Frances Smith.*

---

Manson, M. P. and Grayson, H. M. "The Shipley-Hartford Retreat Scale as a Measure of Intellectual Impairment for Military Prisoners." *Journal of Applied Psychology*, XXXI (1947), 67-81.

The *Shipley-Hartford Retreat Scale* is found, by means of various statistical analyses, to be valid in its use with General Court prisoners as a measure of intelligence only, not as one of deterioration. CQ's obtained from the scores of the prisoners, if taken at face value, evidence some degree of deterioration in the majority of subjects. Factors other than deterioration which may yield deterioration scores are suggested, and it is pointed out that average test scores on the A.G.C.T. indicate that obtained vocabulary scores were unduly high for the prisoner group. It is important that factors affecting the CQ be identified by further experimental research. *Frances Smith.*

---

Marks, Eli S. "Selective Sampling in Psychological Research." *Psychological Bulletin*, XLIV (1947), 267-275.

The author points out that without random selection it is not possible to determine whether the results obtained represent basic characteristics of the population or whether they are by-products of the sampling process. He notes that in many cases sub-populations arise from compound rather than from independent selective factors.

The importance of control of the sampling process is brought out by an analysis of two publications of psychological research. *Leroy S. Burwen.*

---

Meyer, Edith and Simmel, Marianne. "The Psychological Appraisal of Children with Neurological Defects." *Journal of Abnormal and Social Psychology*, XLII (1947), 193-205.

This article illustrates how the inadequacies of prevailing quantitative test methods, when applied to children with severe neurological disorders, can largely be met by a broader theoretical orientation and by other procedures based on genetic studies of reasoning and attention. Modifications in standard examination techniques, together with some less well-known methods, are described, and two case studies of their application are given. The contribution of these supplementary procedures to differential diagnosis and to a better treatment approach is discussed in the concluding section, where three types of handicapped children are distinguished: (1) those with pre- or neo-natal impairment; (2) those with impairment following severe damage after varying periods of normal development; and (3) those with phasic disturbances of mental functioning. *Vernon S. Tracht.*

---

Napoli, Peter J. "Interpretive Aspects of Finger-Painting." *Journal of Psychology*, XXIII (1947), 93-132.

The author suggests guides for interpreting finger-painting productions. These fall into three categories: (1) the administrator's observations of the subject's activity, (2) "the physical manifestations of the patient during the process and of his painting," and (3) the verbalizations of the patient. Among the observations made are the working posture, the placement of paint, the parts of hand used, the type of pattern painted, the patient's interest, and his ultimate satisfaction or dissatisfaction with his production. In the second area are discussed such factors as handedness, colors employed, type of motion made, rhythm, and texture. The findings reported are empirically derived and are now being statistically validated. *Harold Mosak.*

---

Rotter, Julian B. and Willerman, Benjamin. "The Incomplete Sentences Test as a Method of Studying Personality." *Journal of Consulting Psychology*, XI (1947), 43-48.

The incomplete-sentences technique was adapted for use in the Army Air Forces Convalescent Hospitals as a measure of adjustment in terms of psychological fitness for return to duty. The final form of the scale used is described with details of the scoring methods, which were devised in terms of conflict, neutral, and positive test responses. A test reliability of .85 was obtained by the split-half technique, with one psychologist scoring the 200 records used. Inter-scoring reliabilities were also computed for seven scorers on 50 of these records, with an average intercorrelation of .89. Criterion for test

validity was the judgment of the psychologist during the initial interview with each patient. *Frances Smith.*

---

Strang, Ruth. "Methods, Technique, and Instruments of Mental Hygiene Diagnosis and Therapy." *Review of Educational Research*, XVI (1946), 436-443.

Pointing to the increase in serious mental disorders today, the author discusses briefly the progress of the last three years in meeting the need for better psychiatric services. Mention is made of the contributions of experimentation and theory from clinical observation and experience, as found in the literature covering this period. The following topics are treated: clinical application of psychological tests, development of screening devices, the search for syndromes, projective techniques, shock therapy, group therapy, occupational therapy and rehabilitation, counseling and psychotherapy, case studies, and the evaluation of mental hygiene techniques. *Vernon S. Tracht.*

---

Strong, E. K., Jr. "Differences in Interests Among Public Administrators." *Journal of Applied Psychology*, XXXI (1947), 18-38.

After summarizing his previous articles on the interests of public administrators, the author discusses the nature of the differences of interest among public administrators and the degree of these differences. Employing a sample of 552 public administrators who were divided into sixteen functional subgroups, he indicates that public administrators receive no A or B-plus ratings on the thirty-four occupational scales because of the diversity of the subjects' interests. The author concludes with a discussion of the similarities and differences in interests obtaining between the various subgroups. *Harold Mosak.*

---

Thompson, George G. and Witryol, Sam L. "The Relationship Between Intelligence and Motor Learning Ability, as Measured by a High Relief Finger Maze." *Journal of Psychology*, XXII (1946), 237-246.

Forty college students learned a high-relief finger maze to a criterion of three errorless trials. Scores were obtained for each subject in terms of time, errors, and trials. The *Otis Gamma Test* was also administered to these subjects to obtain an estimate of intelligence. Results indicate that the measures of motor learning employed are all highly intercorrelated. While correlations between intelligence and each of these measures are low, the authors attribute this finding to the homogeneity of intelligence in the experimental group. In a more heterogeneous population the authors predict that the correlations would exceed .70 between intelligence and each of the measures of motor learning. *Harold Mosak.*

Vernon, P. E. "An Experiment on the Value of the Film and Film-Strip in the Instruction of Adults." *British Journal of Educational Psychology*, XVI (1946), 149-159.

Seven groups of recruits receiving training in seamanship in the Royal Navy were instructed in different ways and then were given a short, unannounced examination after a week or two had elapsed. The author noted an improvement in examination grades when film strips were introduced, and that the addition of films aided in comprehension rather than in memory for specific details. *Harold Mosak.*

---

Wall, W. D. "Reading Backwardness Among Men in the Army." *British Journal of Educational Psychology*, XVI (1946), 133-148.

Several performance intelligence tests were administered to 100 men in the British Army who had given evidence of reading backwardness. Most of these men were discovered to be in the dull range, but nevertheless educable. In addition, observations were made to detect perceptual errors, speech difficulties, and sensory defects, and a "nervous temperament" test was administered. The author suggests that three characteristics—auditory and visual defects, left-handed dominance, and "nervous temperament"—could differentiate between this group and a normal group described in a previous study. The article concludes with material drawn from case studies. *Harold Mosak.*

---

Wells, F. L. "Verbal Excess Over Quantitation: Two Case Studies." *Journal of Psychology*, XXIII (1947), 65-82.

The author examines the material from two case studies in an effort to relate this evidence to his system of trait-complexes which is described in *Outstanding Traits*. The cases were chosen to illustrate the psychometric patterns of the "Verbalist," "Inarticulate," "Physical Science Motivations," and "Inarticulate" and "Physical Science Motivations." In addition, these cases illustrate the personal histories of subjects who demonstrate a psychometric excess of verbalism over quantitation and of quantitation over verbalism. *Harold Mosak.*

---

Wells, F. L. "Verbal Facility: Positive and Negative Associations." *Journal of Psychology*, XXIII (1947), 3-14.

This paper is a continuation of the author's Grant Study and is based upon the monograph, *Outstanding Traits*. The purpose of the study was to determine the relationship of two verbal traits, "Verbal Facility" and "Inarticulate," to such areas as the social and physical sciences. The data were examined with reference to the Sheldon system of typology. *Harold Mosak.*

Wesman, Alexander G. and Bennett, George K. "The Use of 'None of These' as an Option in Test Construction." *Journal of Educational Psychology*, XXXVII (1946), 541-549.

Of a group of 591 applicants to schools of nursing, 301 took standard multiple-choice tests with options specified; 290 took the test with the fifth option as "none of these" for each item. All were given two twenty-item, five-choice tests: one in Vocabulary and one in Arithmetic. The two groups were found to be comparable on the basis of the Potts-Bennett Vocabulary and Numerical Tests. Item-test coefficients were calculated against two criteria, the test of which the item was a part, and the appropriate Potts-Bennett test. Neither with the Vocabulary Test nor with the Arithmetic Test did the use of "none of these" give greater reliability. Neither item style showed uniform superiority, although the tests with "none of these" items took longer. The authors speculate that the effectiveness of the use of this type of item is perhaps dependent on the quality of the other options. *Leroy S. Burwen.*

---

Wheatley, Luis Andres, and Sumner, F. C. "Measurement of Neurotic Tendency in Negro Students of Music." *Journal of Psychology*, XXII (1946), 247-252.

For purposes of measuring neurotic tendency in Negro music students at Howard University, and of determining its relation to other factors, 81 women and 19 men were given the *Bernreuter Personality Inventory*, the *Otis Self-Administering Tests of Mental Ability*, and the *Sims Score Card for Socio-Economic Status*, while 75 of these subjects were given the *Allport-Vernon Study of Values*. A summary of the findings indicate that these students on the average fall within the so-called normal range; and that the higher neurotic scores correlate positively with lower socio-economic status (100 chances in 100), with higher test intelligence (98 chances in 100), and with more aesthetic-mindedness (90 in 100). The mean for the 75 (mostly women) on the Allport-Vernon conforms to that for white women. *Vernon S. Tracht.*

---

Williams, Meyer. "An Experimental Study of Intellectual Control Under Stress and Associated Rorschach Factors." *Journal of Consulting Psychology*, XI (1947), 21-29.

Attempting experimentally to validate certain components of the Rorschach test with an operational definition of intellectual control, the author discovers that the theoretical assumption that "poor integration of form with color in the Rorschach reflects poor control and non-efficient mental activity in the emotional situations of everyday life" is borne out. The results of the study cannot be generalized, however, because of the small sample of homogeneous, "normal" college students which the author employed. *Harold Mosak.*

Wylie, Ruth C. "Reliability of the Grove Modification of the Kent-Shakow Formboard Series." *Journal of Applied Psychology*, XXXI (1947), 155-159.

The following tests were administered to a group of 352 boys chosen at random from grades 7-12 of the Beaver Falls, Pennsylvania, Public Schools: (1) Grove's modification of the *Kent-Shakow Formboard Series*, (2) Block Designs Test from the *Wechsler-Bellevue Intelligence Scale*, (3) Cube Construction Test from the *Cornell-Coxe Scale*, (4) O'Connor's *Wiggly Block Test*, (5) *Revised Minnesota Paper Formboard, Series AA*, and (6) *California Test of Mental Maturity*. The *Wiggly Block Test* was eliminated because of poor reliability, and split-half reliabilities were computed for the others. It was concluded that the reliability of the Grove Modification of the *Kent-Shakow Series* is equal to or better than the reliability of comparable tests of its type, that further standardization of it is warranted, and that the use of more scores does not seem justified in view of reliability findings. *LeRoy S. Buttern*.

---

#### ADDITIONAL ARTICLES NOT ABSTRACTED

- Anderson, Rose G. "Wimberley's Criticisms of Kuhlmann-Anderson Tests." *Journal of Educational Psychology*, XXXVII (1947), 45-50.
- Baxter, B. "Reliability and Validity of the Kopas Wage Earner Battery of Tests." *Journal of Applied Psychology*, XXXI (1947), 39-43.
- Cross, Orrin H. "Braille Edition of the Minnesota Multiphasic Personality Inventory for Use with the Blind." *Journal of Applied Psychology*, XXXI (1947), 189-198.
- Duncan, Acheson J. "Some Comments on the Army General Classification Test." *Journal of Applied Psychology*, XXXI (1947), 143-149.
- Gough, Harrison G. "Simulated Patterns on the Minnesota Multiphasic Personality Inventory." *Journal of Abnormal and Social Psychology*, XLII (1947), 215-225.
- Graham, Frances K. and Kendall, Barbara S. "Note on the Scoring of the Memory-for-Designs Test." *Journal of Abnormal and Social Psychology*, XLII (1947), 253.
- Harris, Robert E. and Thompson, Clare Wright. "The Relation of Emotional Adjustment to Intellectual Function—A Note." *Psychological Bulletin*, XLIV (1947), 283-287.
- Hogadone, Edwina and Smith, Leo F. "Some Evidence on the Validity of the Cardall Test of Practical Judgment." *Journal of Applied Psychology*, XXXI (1947), 54-56.
- Hunt, William A. "Negro-White Differences in Intelligence in World War II—A Note of Caution." *Journal of Abnormal and Social Psychology*, XLII (1947), 254-255.

- Johnson, A. P. "An Index of Item Validity Providing a Correction for Chance Success." *Psychometrika*, XII (1947), 51-58.
- Jurgensen, C. E. "Table for Determining Phi Coefficients." *Psychometrika*, XII (1947), 17-29.
- Krueger, William C. F. "Influence of Difficulty of Perceptual-Motor Task upon Acceleration of Curves of Learning." *Journal of Educational Psychology*, XXXVIII (1947), 51-53.
- Lawshe, C. H., Jr. and Forster, Max H. "Studies in Projective Techniques: I. The Reliability of a Multiple-Choice Group Rorschach Test." *Journal of Applied Psychology*, XXXI (1947), 199-211.
- Louitt, C. M. and Browne, C. G. "The Use of Psychometric Instruments in Psychological Clinics." *Journal of Consulting Psychology*, XI (1947), 49-54.
- McClelland, David C. "Further Application of Simplified Scoring of Bernreuter Personality Inventories." *Journal of Applied Psychology*, XXXI (1947), 182-188.
- Portenier, Lillian G. "Abilities and Interests of Japanese-American High School Seniors." *Journal of Social Psychology*, XXV (1947), 53-61.
- Smith, Henry P. "The Relationship Between Scores on the Bell Adjustment Inventory and Participation in Extra-Curricular Activities." *Journal of Educational Psychology*, XXXVIII (1947), 11-16.
- Voss, Harold A. "Analysis in Terms of Frequencies of Differences." *Psychometrika*, XII (1947), 43-50.
- Wesman, Alexander G. "Active Versus Blank Responses to Multiple-Choice Items." *Journal of Educational Psychology*, XXXVIII (1947), 89-95.



## THE CONTRIBUTORS

Walter W. Cook—Ph.D., University of Iowa, 1931. Professor and Director of Teacher Training, Eastern Illinois State Teachers College, 1931-1938. Visiting Professor, University of Chicago, summers of 1932, 1933, 1936. Visiting Professor, University of Pennsylvania, summer of 1938. Professor of Education, University of Minnesota, 1938-. Author of articles and monographs dealing with technical aspects of test construction, individual and trait differences, promotion and grouping policies in the elementary schools. Member, American Psychological Association, American Educational Research Association, American Association of School Administrators, Phi Delta Kappa, Kappa Delta Pi, Sigma Phi Epsilon, Psi Chi, Pi Kappa Delta.

David L. Ekberg—B.S., Northwestern University, 1940. Personnel Test Technician, Standard Register Company, 1940-1941. Officer in U. S. Army, 1942-1945, Adjutant General's Department; Classification Officer, Army Air Bases in U. S. and China-Burma-India Theater; Wing Personnel Officer, China Wing, Air Transport Command, Kunming, China, 1944-1945. Assistant to the Director of Employee Relations, Standard Register Company, 1946. Personnel Test Technician, Air Service Command and Air Material Command Headquarters, 1942, 1946. Author of tests of mechanical and clerical aptitudes. Member, National Vocational Guidance Association.

Max D. Engelhart—Ph.D., University of Illinois, 1932. Instructor in Chemistry, Georgia School of Technology, 1926-1927. Assistant, Bureau of Educational Research, University of Illinois, 1927-1932. Research, University of Chicago, 1933-1934. Director, Department of Examinations, Chicago City Junior College, 1935-. Member of Examination Staff, U. S. Armed Forces Institute, 1943-1944. Co-author of *Scientific Study of Educational Problems* and the *Chicago Reading Tests*. Author of articles on measurement, statistical method, and techniques of educational research.

Clifford P. Froehlich—M.A., University of Minnesota, 1939; graduate work, George Washington University, 1946-. State Supervisor, Occupational Information and Guidance Service of North Dakota, 1940-1942. With the U. S. Army, Psychological Research Unit of the Army Air Corps and Clinical Psychology Program, 1942-1946. Specialist, Occupational Information and Guidance Service, U. S. Office of Education, 1946-.

Richard H. Gaylord—M.A., University of Florida, 1941. Instructor in Psychology, 1941-1942; Associate Director, University Testing Service, 1942-1945, University of North Carolina. Personnel Technician, Personnel Research Section, AGO, War Department, 1945. Instructor in Psychology, Ohio State University, 1945-1946. Assistant Chief, Statistical Analysis Unit, Personnel Research Section, AGO, War Department, 1946-. Member, American Psychological Association, Psychometric Society.

Robert John Keller—Ph.D., University of Minnesota, 1947. Teacher, elementary and high school, White Bear Beach and North St. Paul, Minnesota, 1931-1940. Teaching and Research Assistant, University of Minnesota, 1940-1942. With the U. S. Army: Psychological Assistant, Psychological Research Unit No. 2, SAACC, Aug. 1942-Jan. 1945; Clinical Psychologist, AAF Convalescent Hospital, Jan. 1945-Oct. 1945; Aviation Psychologist, Psychological Branch, June 1945-Oct. 1945. Teaching and Research Assistant, University of Minnesota, 1945-1946. Visiting Professor of Secondary Education, College of Education, University of Minnesota, summer 1947. Research Psychologist, Psychological Research and Indoctrination Division, Air Training Command, San Antonio, Texas, 1946-. Member, American Psychological Association, Psychometric Society, National Society for the Study of Education, Phi Delta Kappa, Kappa Delta Pi.

J. A. Morris Kimber—Ph.D., University of Southern California, 1945. Professor of Biblical Literature and Instructor in Psychology, Whittier College, California, 1916-1919. Harrison fellow in Psychology, University of Pennsylvania, 1919-1920. Graduate student, University of California, 1920. Graduate student, Stanford University, 1922. Registrar and Professor of English and Psychology, Columbia Bible College, South Carolina, 1925-1943. Lecturer in Psychology, University of Southern California, 1945. Senior Counselor, Veterans Guidance Center, University of Southern California, 1945-1946. Assistant to the Dean of University College in charge of Extension Division, and Lecturer in Psychology, University of Southern California, 1946-. Associate member, American Psychological Association. Member, Phi Delta Kappa.

Carroll H. Leeds—Ph.D., University of Minnesota, 1946. Teaching and Research Assistant, University of Minnesota, 1939-1941. Head of the Department of Education and Psychology and Director of Teacher Training and Placement, Westminster College, New Wilmington, Pa., 1941-1946. Head of the Department of Psychology, Mississippi State College for Women, 1946-. Associate member, American Psychological Association. Member, Phi Delta Kappa, Phi Chi.

Milton M. Mandell—B.A., New York University, 1933. Assistant Director of Examinations, Los Angeles City Civil Service Com-

mission, 1939-1940. Classification Consultant, State of Connecticut, 1940-1941. Regional Personnel Officer, OEM, 1941-1942. Personnel Officer, Office of Program Vice-Chairman, War Production Board, 1942-1943. Chief Analyst, Committee for Congested Areas, 1943-1944. Chief, Administrative and Management Testing, U. S. Civil Service Commission, 1944-. Member, American Society of Public Administration, Civil Service Assembly.

**Donald E. Super**—Ph.D., Columbia University, 1940. Assistant Employment Secretary, Y.M.C.A., Cleveland, Ohio, 1932-1935. Director, Cleveland Guidance Service, Cleveland, Ohio, 1935-1936. Research Assistant, Teachers College, Columbia University, 1936-1938. Associate Professor of Educational Psychology and Director, Student Personnel Bureau, Clark University, 1938-1942. Aviation Psychologist, Army Air Forces, 1942-1945. Associate Professor of Education, Department of Guidance, Teachers College, Columbia University, 1945-. Author of articles in psychological and educational journals and of aviation psychology research reports. Author of two books on measurement and guidance. Fellow, American Psychological Association (President-elect, Division of Consulting Psychology). Member, New York State Association of Applied Psychologists, Eastern Psychological Association, National Vocational Guidance Association, and American College Personnel Association.

**Erwin K. Taylor**—Ph.D., Northwestern University, 1941. Personnel Examiner, Illinois State Civil Service Commission, 1942-1943. Personnel Technician, Personnel Research Section, Adjutant General's Office, 1943-1945. Chief, Statistical Analysis Unit Personnel Research, Adjutant General's Office, 1945-. Fellow, American Psychological Association. Member, Psychometric Society, Civil Service Assembly of U. S. and Canada.



# TABLE of CONTENTS<sup>1</sup>

## PROCEEDINGS

### AMERICAN COLLEGE PERSONNEL ASSOCIATION

MARCH 29

#### GENERAL SESSION 9:15 A. M.

Presiding, WILLARD W. BLAESSER, Director, Student Personnel Service, University of Montana

Presidential Address: *When Colleges Bulge*

D. D. FEDER, Dean of Students, University of Denver ... 475

*Can Colleges and Universities Contribute to Industrial Peace*

DONALD J. SHANK, Director of Student Personnel, New York State School of Industrial and Labor Relations, Cornell University (To be published in *Journal of the Association of Collegiate Registrars*)

#### BUSINESS MEETING 11:00 A. M.

Presiding, DANIEL D. FEDER, Dean of Students, University of Denver, and President, ACPA

*Journal of Educational and Psychological Measurement*, FRED MCKINNEY, Professor of Psychology, University of Missouri

ACPA Membership, THELMA MILLS, Director of Student Affairs for Women, University of Missouri, and Secretary, ACPA

#### LUNCHEON MEETING 12:15 P. M.

Presiding, C. GILBERT WRENN, Professor of Education, University of Minnesota

*Graduate Record Examinations*

K. W. VAUGHN, Director, The Graduate Record Examination (To be published in *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*)

#### GENERAL SESSION 2:30 P. M.

Presiding, HELEN G. FISK, Associate Director, Western Personnel Institute

*A Review of New Books and Pamphlets on Occupations for College Students (1942-1946)*

ROBERT HOPPOCK, Professor of Education, New York University ..... 485

*Symposium: Contributions of Military Testing to Educational and Industrial Personnel Work*, DEWEY B. STUIT, Chairman

<sup>1</sup> ACPA Program Committee, Chairman, Wendell S. Dysinger, Dean of the College and Director of Personnel, MacMurray College.

ACPA Convention Committee of Recorders, Chairman, Corinne LaBarre, Research Assistant, Western Personnel Institute.

ACPA Committee on Convention Arrangements, Chairman, Francis P. Robinson Professor of Psychology, Ohio State University.

*The Effect of the Nature of the Criterion Upon the Validity of Aptitude Tests*

DEWEY B. STUIT, Director, Student Counseling Office and Professor of Psychology, State University of Iowa  
(To be published in EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT)

*Gadet Personnel Problems and Procedures at the United States Military Academy*

T. ERNEST NEWLAND, Associate Director, Department of Military Psychology and Leadership, United States Military Academy ..... 489

*Importance of Criteria in Selection and Training*

A. C. VAN DUSEN, Associate Professor of Psychology and Director, University College Bureau of Counseling, Northwestern University ..... 498

*Defects in Educational and Personnel Systems as Reflected in Military Experiences*

WILBUR S. GREGORY, Head, Counseling and Testing Services, University of New Mexico ..... 505

*The Validity of Standard and Custom-Built Personality Inventories in a Pilot's Selection Program*

DONALD E. SUPER, Associate Professor of Education, Teachers College, Columbia University (To be published in EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT)

*What the Veteran Thinks of Education*

HAROLD M. MORRIS, Graduate Student, New York University ..... 512

**SOCIAL HOUR 4:30 P. M.**

Chairman, DOROTHY V. N. BROOKS, Dean of Women, Denison University

**BANQUET 6:30 P. M.**

Presiding, FORREST H. KIRKPATRICK, Dean of Students, Bethany College, and Educational Counselor, Radio Corporation of America

*Impressions of Student Personnel Work in German Universities and Implications for America*

E. G. WILLIAMSON, Dean of Students, University of Minnesota ..... 519

**GENERAL SESSION 8:30 P. M.**

Presiding, HELEN M. VOORHEES, Director, Appointment Bureau, Mount Holyoke College

*Graduate Training for Educational Personnel Work*

CORINNE LA BARRE, Research Assistant, Western Personnel Institute ..... 538

*Graduate Training in Guidance and Counseling*

HUGH M. BELL, Dean, Student Personnel and Guidance, Chico State College (To be published in *Occupations*)

*Future Enrollments in U. S. Institutions of Higher Education*  
 J. H. ROHRER, Associate Professor of Psychology and Research Assistant to the President, University of Oklahoma 547

## MARCH 30

### GENERAL SESSION 9:00 A. M.

Presiding, DANIEL D. FEDER, Dean of Students, University of Denver and President, American College Personnel Association

#### *The College Community and Occupational Integration of Minorities*

FRANK S. LOESCHER, Placement Service, American Friends Service Committee ..... 553

#### Group A

Presiding, RICHARD C. CROSBY, Director of Student Counseling, Miami University

#### *Gradations of Directiveness in Counseling*

EDWARD S. JONES, Dean of Students, University of Buffalo ..... 559

#### *Faculty Counseling for Freshmen*

LONZO JONES, Director, Student Personnel Services, Indiana State Teachers College ..... 564

#### *Discipline: A Student Counseling Approach*

JOHN D. FOLEY, Assistant to the Dean of Students, University of Minnesota ..... 569

#### Group B

Presiding, FRANCES M. CAMP, Director, Educational Placement Office, University of Iowa

#### *The Application of Personnel Methods to University Housing Procedures*

B. JAMES BORRESON, Director, Student Housing Bureau, University of Minnesota ..... 583

#### *Measuring Outcomes of Classroom Therapy*

HAROLD B. PEPINSKY, Assistant Professor, The Basic College Department of Effective Living, Michigan State College (To be published in EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT)

#### *Relationship of the College Administration with Fraternities and Sororities*

GEORGE S. BEERY, Dean of Students, Drake University 594

### GENERAL SESSION 2:00 P. M.

Presiding, ANNA L. ROSE HAWKES, Dean of Students, Mills College

#### *The College Personnel Worker's Responsibility for the Improvement of Reading*

RUTH STRANG, Professor of Education, Teachers College, Columbia University ..... 603

<i>College and University Placement of Graduates in Business</i> KARL W. ONTHANK, Dean of Personnel Administration, University of Oregon (To be published in <i>Occupations</i> )	
<i>School Marks and Success in Engineering</i> GEORGE A. PIERSON, Dean of Students and Professor of Education, University of Utah	612
<i>The Graduate Record Examination vs. Other Measures of Aptitude and Achievement</i> JOSEPH C. HESTON, Director, Bureau of Testing and Re- search, DePauw University	618
<i>Vocational Problems of Undergraduate University Women</i> WILMA T. DONAHUE, Director, Bureau of Psychological Services, Institute for Human Adjustment, University of Michigan (Published jointly with WILMA ELDERSVELD in <i>Journal of Higher Education</i> , XVIII (1947), 194-200)	
<i>Trends and Development of the Vocational and Other Inter- ests of Veterans at Washington University</i> NATHAN KOHN, JR., Veterans' Counselor, Washington University	631
VESPER SERVICE (Cooperating with NADW) 4:30 P. M. Trinity Episcopal Church, East Broad & South Third Sts., Columbus, Ohio	
GENERAL SESSION 7:30 P. M.	
Presiding, LYSLE W. CROFT, Director of Personnel, University of Kentucky	
<i>The Diagnosis of Reading Deficiencies as an Aid to Remedial Work</i> FRANCES ORALIND TRIGGS, Chairman, Committee on Diag- nostic Reading Tests	638
<i>Symposium: "What's New in Personnel Procedures?"</i> Leader: ESTHER LLOYD-JONES, Associate Director of Personnel, Teachers College, Columbia University. Recorder: WIL- LIAM M. GILBERT, Acting Director of the Student Per- sonnel Bureau and Assistant Professor of Psychology, University of Illinois	
<i>An In-Service Training Program for Residence Counselors</i> DUANE COLLINS, Assistant Professor of Education, Uni- versity of Connecticut	647
<i>The Use of Dormitories for Social Education</i> FLORENCE THOMPSON, Dean of Women, Michigan State College	648
<i>Coordination of Student Personnel Services at Cornell Uni- versity</i> LUCILLE ALLEN, Counselor, Cornell University	655
<i>Current Trends in Personnel Work at Pratt Institute</i> MARGARET STEPHENSON, Director of Student Life, Pratt Institute	657
<i>The National Student Movement and Conference</i> JOHN BERGSTRESSER, Assistant Dean of Students, Uni- versity of Chicago	662
<i>The Keuka College Field Work Plan</i> KATHERINE BLYLEY, President, Keuka College	665



## Presidential Address WHEN COLLEGES BULGE

D. D. FEDER

Dean of Students, University of Denver

FOR almost ten years we have been sighing nostalgically for "the good old days" of peace and quiet, of prosperity and depressions, of civilized warfare without atom bombs and threats of disease invasions, of college campuses returned to the leisurely pursuit of "culture and learning." Now, all too roughly, we are being reminded that we can never go back, that we can never recapture those supposed delights of youth. Instead we are facing the necessity of an immediate and reasonably adequate adjustment to things as they are. And "things as they are" constitute a kaleidoscopic pattern of bulging walls, crowded dormitories, new and previously unthought-of curricula, and, above all, a new and different kind of student body.

It is, of course, logical to assume that the problems encountered on one campus are common to almost every other in the country. Variations will certainly be more matters of degree rather than of kind. It is our purpose here to examine briefly:

1. The nature of present-day college populations.
2. Some of the special characteristics of the veteran population in terms of their needs and motivation in coming to college.
3. The generally changing motivation and orientation of all college students.
4. The need for high caliber professional services in vocational, educational, and personal counseling for all students.
5. Some problems of the classroom being brought to light through counseling interviews with especial reference to the military treatment of similar situations.

6. Some ways in which the integrated personnel service program may serve both faculty and student body in more effectively meeting student needs.

The common pattern of higher education in America today is a student body one-half to two-thirds veteran. The remainder is composed of the typical recent high-school graduates to whom college has traditionally meant glamour, athletics, "Greek" life, manifold extra-curricular activities, and occasional study. In stark contrast is the majority of veterans who after three to five bitter years have returned to the hope and promise held out to them by the opportunity for self-improvement through education. For them—not 100 per cent of them to be sure, but the majority—there is a serious purpose and hope for a better life in all ways. For them there is hope for the opportunity to rise above their present station and to partake of a fuller life, socially, culturally, and perhaps primarily, economically.

It is reasonable to ask whether college education has been oversold to the returning G.I. From every side he has been bombarded with glowing statements as to the great service his country is rendering him in making available to him the "advantages" of a college education in our modern, machine economy. There have been numerous statements, directly and by implication, that he could earn a more plentiful living if he had a college degree. By the same token he has been assured from many directions that he would find in his college courses answers to the many "whys" he asked of himself and his buddies as he slogged through the mud and horror of war. Why wars? Why depressions? Why unemployment? Why bigotry and intolerance? Why unhappiness in the midst of plenty? Why starvation of soul in the midst of the glutting of body? All these and many more he has been told would fall into orderly procession and be answered as he fed on the diet of education which would prepare him for living as a free man in a free world.

It is no exaggeration to say that the veterans have returned to American campuses with great expectations. It is similarly no exaggeration to say that a significantly large number of

veterans have, in the course of even a few weeks or months, experienced great disappointments. There is a very real danger that unless immediate and realistic remedial measures are adopted, these same disappointments may become multiplied and communicated to the non-veteran population as well, with the result that higher education may suddenly find itself relegated to the role of preparing only for the higher professions with their "trade union" requirements of pre-professional educational patterns.

In the time available we shall be able to examine but a few of the adjustments being made or which need to be made by higher education in the attempt to serve our country significantly in this newest crisis of peace. Actually many of these adjustments have been in process on many campuses in the last two decades and are now simply being accelerated by the "veterans' bulge" which in itself may be a passing phenomenon. But the resultant events and adjustments are fully as significant for the non-veteran since both groups are being assimilated into a unitary student body on most campuses.

A primary and major adjustment stems from the enthusiastic overcrowding of physical facilities. The result—bulging classrooms, overpopulated and inadequate dormitories, long lines waiting for every type of service. A few—but very few—universities have anticipated this situation by the erection of temporary classrooms, dormitories and other building facilities. To such schools the "veterans' bulge" does not represent a source of profit since the school must make expensive outlays in its effort to serve student needs. In marked contrast are the schools content to crowd sixty students into classrooms designed for thirty.

A corollary to the overcrowded classroom is the overloaded instructor. Where classes of twenty-five to thirty have given way to classes of fifty to several hundred, instruction has turned from an intimate attention to individual needs to the impersonal methods of the production line. Not only does the instructor lose sight of the student as an individual, but he must perforce become so immersed in the routines of class accounting as to find little or no time for the art of teaching.

Let it not be assumed that students will accept these conditions good-naturedly and unquestioningly. Neither veteran nor non-veteran anticipated assembly-line operations in the classroom and hence he is frankly disappointed and dissatisfied with them. Of a group of transferees interviewed shortly after the opening of the fall term, a number gave as their primary reason for coming to the University the fact that it had kept its classes small by adding both room and faculty. Some colleges may pride themselves on excellent bank balances as a result of overcrowding classrooms and overloading faculty, but this is short-range and short-sighted economy. Students may accept such treatment when, in desperation, they cannot go elsewhere, but such practice smacks of profiteering and racketeering and already not only the students themselves, but even the press is beginning to be vocal in their criticism. The institution which cannot freely open its books and point with pride to its record may soon find itself in a state of disrepute far more serious to its future than its failure to play good football. Indeed, all higher education may suffer from the criticism rightfully accruing to only a few institutions.

There have been so many commentaries on the changes in college populations that the points need no further belaboring here. It is important to note, however, that where the veteran population equals or exceeds that of the non-veteran it is quite likely that the former by virtue of maturity, and being more vocal and generally more aggressive, will tend to impose the pattern of their desires upon the younger non-veteran group.

It is not to be expected that a man with a wife and child will be inclined to participate with much enthusiasm in adolescent horseplay. Similarly some of the trivia which pass for traditions on many campuses may be thoroughly irksome to the intellectually mature non-veteran as well as to the veteran. Therefore, in extra-class relationships we will do well to examine what the student body has traditionally required of itself. Guidance in the direction of worthwhile traditions, of activities with desirable goal reference is an essential function of the Student Personnel Service program.

Colleges are used to youthful high-school graduates with

their unquestioning enthusiasm and respect for authority. Veterans who gave up their status as free-thinking, free-speaking citizens during the war years of stringent discipline feel differently, however, on their return to civilian life. Even the military recognized the importance of allowing men to "blow their tops." The Army provided the "B-Bag" in *Stars and Stripes* in which a man might air his gripes on any subject without fear of reprisal and generally with the knowledge that something constructive might result if his case were legitimate. The student personnel services should be free of any suspicions of reprisal in order to provide the needed safety valve. Such contacts may well be the opening wedge for the counseling of many students who might not recognize their personal needs and might not otherwise seek the services available to them.

From their experiences with military training veterans bring to the college campus certain attitudes and requirements which, if we take proper heed, may lead us to practices in higher education which the military found efficient. Although admittedly borrowed in the first place from civilian educational research, the blunt fact is that higher education has gone its way unintentionally or otherwise uncognizant of and unwilling to recognize some of the findings which in the last few pre-war years became almost routine to the counselor.

Present-day students, both veteran and non-veteran, seek a specific goal orientation for their educational program. Educational dilettantes exist in both groups, but for most of them educational objectives are real and earnest. Recent surveys of the interests of returning G.I.'s show a predominance of vocational orientation even among those enrolled in liberal arts colleges. This should be satisfying rather than alarming to educational classicists. Education during the Classical and Renaissance periods was designed to endow free men and courtiers with the skills, attitudes, knowledge and techniques necessary for earning a livelihood as well as for the art of effective living. As modern education turns toward the equipping of our students with the techniques of effective living in our times it is actually turning in the direction of the classical tradition set to the tune of electronics, atom bombs, and jet speeds.

Demobilization counseling and the widely advertised counseling services of the Veterans Administration have given an increased impetus to all vocational counseling services. When given the opportunity to receive such services recently, a large number of veterans attending school under P.L. 346 immediately volunteered for counseling; as several expressed it "just to be sure." These men are seeking not only the assurance that they are working toward vocational goals appropriate to their interests and abilities, but also that such goals represent adequate opportunities for making a living. Therefore, in our counseling services the case is not closed until the record shows among other things, that the student has consulted the best, most up-to-date sources of information about the status of the vocations in which he may have legitimate interests. No counselor can possibly keep himself informed of developments in the myriad careers in business and industry today. Therefore, accurate, timely information services become a crucial tool to the counselor and his clients. A first step in job satisfaction is a realistic knowledge of what to expect in the field.

A second important demand of students is their desire to know that there is a specific useful purpose in each course they take. This stems in large part, certainly, from military training with the specific purposeful goal for each day of instruction. Returning to civilian pursuits, veterans find lack of personal security as well as lack of motivation for study in courses whose goal in their over-all educational pattern is not clear. This need emphasizes the importance of "selling" courses on legitimate grounds rather than as pre-requisites or requisites. The opening days of instruction should be devoted to clearly stating the purpose and function of the specific course in the student's program.

Stemming directly from the desire to be convinced of the purpose of each course, there is a correlative desire by the student for objective goals of instruction. Both secondary and military education have outstripped higher education in their definition of the specific objectives of instruction in a given course. The sailor who took a course of instruction in the operation and maintenance of radar equipment knew from the

outset what skills and knowledge were his goals. Returned to the college campus he finds the vagueness of objectives in most of his courses a source of insecurity and hence irritation. The remedy lies in the regular practice of requiring instructors to set down in accurately described, achievable terms, definitions of the skills, knowledge and attitudes which are the expected end-products of instruction. This requirement is not an invasion of academic freedom, but an essential part of fulfilling our contract with our students.

Already many campuses are experiencing the fourth requirement of Joe College '46—the need for more realistic instruction. In exit interviews mature veterans with high scholastic ability indicate that they are withdrawing completely from further education because of their disappointment at the meagerness of the fare offered. They want meat, but we all too often give them milk and toast. Others have no hesitancy in requesting course changes because of the inconsequential nature of the subject-matter being presented. Others feel, with justice, that their maturity entitles them to treatment above the level of the secondary school. Others have no hesitancy in absenting themselves from a class in which they feel “nothing is going on.”

Military education soon learned the importance of duplicating as completely as possible, even down to the element of actual danger, the combat situation. Of any course we may legitimately ask—How realistic is instruction? How close to current life situations do discussions of economics, political science, and sociology come? How often does college instruction strike out beyond the lecture and discussion to explore the fields of experience, not merely as chance supplements to, but as more valid content than the pallid fare of the printed page? When American colleges converted to specialized training for military purposes they strove mightily for such realism. There is a disquieting unwillingness to strive in our program of education for peace with energy equal to that we expended in training for the arts and sciences of war.

Out of the psychological laboratory a good many years ago came the information that specific and regular knowledge of his own status (technically called “knowledge of results”) resulted

in increased efficiency on the part of the learner. The armed forces capitalized on this laboratory finding by providing each man almost daily with specific grades indicating his progress in the training program. With the values of such experience still fresh in mind veterans seek such information on at least a weekly basis rather than leaving it a secret until the end of the course. The extra effort required of the instructor who provides weekly reports of status to his classes will be rewarded by significantly improved learning efficiency on the part of his students.

Under the impact of the "veterans' bulge" colleges have found themselves willingly or unwittingly providing a whole pattern of new services to their students. The pre-war trend toward the grouping of all such student services into a single personnel service is growing nationally. The extent of such services is determined, of course, by local policy. Certain minima are required, however, if we are to meet the expressed needs of students. They are:

1. The provision of high caliber professional guidance services. These will include the use of faculty as well as full-time professional counselors and will offer the student aid in vocational, educational, and personal problem areas.

2. The provision of up-to-date vocational information so that any time during his preparatory career, the student may assay his status in the rapidly changing job scene.

3. The provision of continuous counseling services to assist the veteran in passing through the successive stages of change from authoritarian submissiveness to civilian independence and responsibility, and similarly to assist the non-veteran in his quest for maturity.

4. The provision of comprehensive, authoritative information services to avoid the bane of military existence—the passing of the buck.

Where the college has many services under separate heads or where new services have been added recently without regard to their correlation with existing facilities there are numerous instances of conflicting counseling being rendered, to the confusion and disgust of the students. The veteran who is receiving psychiatric service through the VA, the school psychiatrist,



and other incidental counseling, may be overcounseled to the point of unhealthful dependence or overstimulated to unwarranted confusion. Too many "cooks" may not only negate each other's efforts, but create new conflicts which had no prior existence. Experiences such as these emphasize the striking need for coordination and cross-checking of a student's contact with various services through a central clearing house.

Failure to detail the many services which make up a personnel program, because of lack of time, should not lead to the inference that the coordinated personnel service is to be confined only to vocational and educational counseling. In the ideal situation, all extra-class services will be members of the family of student personnel services, with a central clearing station, a central record, and staff clinics designed to bring about through cooperative discussion, analysis and action the optimum service to the individual.

Through close liaison with faculty and administration the student personnel services may serve the student body in yet other ways:

1. By transmitting to the appropriate authority the information concerning expressed student needs, personnel services may influence both curriculum and instruction in the direction of greater service to the student.

2. By acting as a sounding board for hearing and transmitting student problems, needs, and gripes, personnel services have a very direct function in aiding the colleges to render more efficient service.

In appraising the total picture of higher education, there is much satisfaction to be gained from the foresightful programs of counseling services which have developed during the last two decades. The adoption of many techniques by our armed forces lends extra credence to our efforts.

To the end that we strive to fit the individual for adjustment to his contemporary world, and in so doing fit education to the needs of the individual, student personnel functions are truly in the spirit of classical education for effective living.

The field of student personnel work has suffered from an ill of its own making, namely a tendency to divorce its findings and activities from those of the classroom. Our responsibilities

are not fully discharged merely through our service to the individual in an office-counseling situation. We must, as a matter of routine, transmit to the instructional staff those findings regarding student reactions and needs which will assist the faculty in the improvement of their instruction by the adjustment of content, procedure and objectives to infuse meaning and purpose into those classrooms which have become sterile through a long-continued dissociation from the realities of contemporary life. The head-on attack is certainly not indicated, but slowly and surely we are obliged to infuse into instruction through individual contact, faculty meetings, and the other available media, the conception of instruction suited to individual needs, of content made vital by its contact with student realities, and a general atmosphere in which the methods and content of instruction are centered upon the student and interpreted in the light of his needs, interests and abilities. Future developments call for the careful appraisal and evaluation of services rendered. Through tradition and seniority some activities may be stronger budgetwise than they deserve to be considering the services rendered. Cost analysis and evaluation of effectiveness are essential in the justification of budget askings. Throughout we must emphasize the elemental fact of the individualization and personalization of student personnel services. The investment in student personnel services is not only for the purpose of delivering the individual to the classroom in optimum condition for learning, but also to deliver him to the community and to the nation for optimum participation in constructive citizenship. The strength of our nation lies in the strength and adjustment of its individual members.

Concerned as we are with the "veterans' bulge" we are likely to assume that the problems which we face currently are exclusively and peculiarly limited to the returned G.I. But these needs are symptoms of the times and what we consider to be good instructional or personnel practices for veterans must be good for all students.

This is no temporary challenge we face today. It is the challenge for higher education's entire future. This is our opportunity to blaze new trails and to strive for new and more vital goals as we serve the needs of our future student generations.

# A REVIEW OF NEW BOOKS AND PAMPHLETS ON OCCUPATIONS FOR COLLEGE STUDENTS (1942-1946)

ROBERT HOPPOCK

Professor of Education, New York University

## *General Comments*

FROM 1942 to 1944 most of the new occupational literature dealt with opportunities in military service and related occupations.

In 1945 and 1946 appeared quantities of new books and pamphlets on postwar occupations for veterans. Some were excellent; many appear to have been prepared hastily for a ready market.

The highest average quality is in the pamphlet series of the principal publishers listed below.

For the first time in years, the opportunities in small business ownership are receiving considerable attention.

The two most significant publications of this period are the books by Shartle and by Forrester, listed below.

## *Cautions*

Beware of recruiting literature published by schools, colleges, professional associations, and military services. Almost invariably it exaggerates the attractions and ignores the disadvantages of the occupation.

Beware of articles in popular magazines by unknown authors who are frequently not meticulous about verifying statements which make good copy.

Go through your library of occupational books and pamphlets and move to the historical section all those more than ten years old.

*Principal Publishers of Occupational Pamphlets  
1942-1946*

Write to each for the latest price list

- Bellman Publishing Co., 6 Park St., Boston 8, Mass.  
 Institute for Research, 537 S. Dearborn St., Chicago 5, Ill.  
 Occupational Index, Inc., New York University, New York 3, N. Y.  
 Science Research Associates, 228 S. Wabash Ave., Chicago 4, Ill.  
 U. S. Dept. of Commerce, Bureau of Foreign and Domestic Commerce, Industrial (Small Business) Series, Washington 25, D. C.  
 U. S. Dept. of Labor, Washington 25, D. C.  
 National Roster of Scientific and Professional Personnel, Bureau of Labor Statistics, Occupational Outlook Division, Women's Bureau.  
 Western Personnel Service, 30 N. Raymond Ave., Pasadena 1, Calif.

*Some of the Better Books and Pamphlets  
from Other Publishers  
1942-1946*

Selected from the *Occupational Index*

- Wessels, O. Richard. *Small business as a career*. 1946. Syracuse University Press, Syracuse, N. Y. 200 pages. \$2.50.  
*Public health: a career with a future*. No date. American Public Health Association, 1790 Broadway, New York 19, N. Y. 29 pages. 10¢.  
*The social worker*. No date. American Association of Schools of Social Work, 1313 E. 60 St., Chicago 37, Ill. 12 pages. Free.  
*Public accountancy as a career*. 1945. American Institute of Accountants, 13 E. 41 St., New York 17, N. Y. 24 pages. Up to 5 copies, free; additional copies, 5¢ each.  
*Check list for establishing a retail business*. No date. Inquiry Reference Service, Bureau of Foreign and Domestic Commerce, U. S. Department of Commerce, Washington 25, D. C. 13 pages. Free.  
*Architecture, a profession and a career*. 1945. The Ameri-

can Institute of Architects, The Octagon, Washington, D. C. 57 pages. 25¢.

*Getting into the trucking business.* 1945. American Trucking Associations, Inc., 1424 16 St., N.W., Washington 6, D. C. 63 pages. Free.

*Shall I take up farming?* 1945. Superintendent of Documents, Washington 25, D. C. 50 pages. 15¢.

Greenleaf, W. J. *Pharmacy.* Guidance leaflet No. 14. Revised. 1945. Superintendent of Documents, Washington 25, D. C. 20 pages. 10¢.

Allen, S. W. *Handbook of information on entering positions in forestry.* 1945. Society of American Foresters, 825 Mills Bldg., 17 St. and Pennsylvania Ave., N. W., Washington 6, D. C. 56 pages. 25¢.

Spero, S. D. *Government jobs and how to get them.* 1945. J. B. Lippincott Co., 521 Fifth Ave., New York, N. Y. 358 pages. \$2.95.

Greenberg, D. B. and Schindall, H. *A small store and independence.* 1945. Greenberg: Publisher, 400 Madison Ave., New York 17, N. Y. 243 pages. \$2.00.

Savord, R. *Special librarianship as a career.* 1945. Institute of Women's Professional Relations, Research Headquarters, Connecticut College, New London, Conn. 16 pages. 15¢.

Noblette, C. B. *If you are considering photography.* Revised edition. 1944. Rochester Athenaeum & Mechanics Institute, Rochester 8, N. Y. 31 pages. 10¢.

*The Federal Bureau of Investigation, U. S. Department of Justice.* 1944. Federal Bureau of Investigation, U. S. Department of Justice, 1435-37 K St., N.W., Washington 25, D. C. 44 pages. Free.

Hinkel, R. E. and Baron, L. *An educational guide in air transportation.* 1944. Transcontinental & Western Airlines, Inc., Kansas City, Mo. 140 pages. 75¢.

Van Peursem, R. L. *If you are considering industrial chemistry.* Vocational Guidance Series, Pamphlet No. 3. 1944. Rochester Athenaeum & Mechanics Institute, Rochester 8, N. Y. 19 pages. 10¢.

Purvis, E. G. *Secretaryship as a career field*. 1944. National Council of Business Schools, Washington 6, D. C. 23 pages. 10¢.

Johnson, H. *Your career in music*. 1944. E. P. Dutton & Co., New York 10, N. Y. 319 pages. \$3.00.

Hurd, L. M. *Modern poultry farming*. 1944. Macmillan Co., New York, N. Y. 599 pages. \$4.00.

Woodhouse, C. G. *The big store*. 1943. Funk & Wagnalls Co., New York, N. Y. 196 pages. \$1.50.

Evans, E. K. *So you're going to teach*. 1943. Julius Rosenwald Fund, 4901 Ellis Ave., Chicago, Ill. 52 pages. 25¢.

Reck, Franklin M. *Radio from start to finish*. 1942. Thomas Y. Crowell Co., New York, N. Y. 160 pages. \$2.00.

*The American Foreign Service*. 1941. Government Printing Office, Washington 25, D. C. 148 pages. Free.

Stewart, Lowell O. *Career in engineering*. 1941. Iowa State College Press, Ames, Iowa. 87 pages. 75¢.

*Dentistry as a professional career*. 1941. Council on Dental Education, American Dental Association, 212 E. Superior St., Chicago, Ill. 72 pages. Free.

#### *Basic References on Occupational Information 1942-1946*

Shartle, C. L. *Occupational Information*. 1946. Prentice-Hall, Inc., 70 Fifth Ave., New York, N. Y. 339 pages. \$4.60—Techniques of occupational research used by U. S. Employment Service and others. Best single reference on how to collect facts about jobs from original sources.

Forrester, G. *Occupations: a selected list of pamphlets*. 1946. The H. W. Wilson Co., 950 University Ave., New York, N. Y. 238 pages. \$2.50—Annotated list of 1500 pamphlets grouped by occupation. Principal publishers. How to select, evaluate, and file occupational pamphlets.

*Occupational Index*. Occupational Index, Inc., New York University, New York 3, N. Y. \$5.00 a year—Quarterly bibliography of new books and pamphlets on all occupations. Annotated and indexed by occupation. Recommended references starred in margin.

# CADET PERSONNEL PROBLEMS AND PROCEDURES AT THE UNITED STATES MILITARY ACADEMY<sup>1</sup>

T. ERNEST NEWLAND

*Associate Director, Department of Military Psychology and Leadership,  
United States Military Academy*

THE specificness of the educational and training commitment of the United States Military Academy is much more marked than in the cases of most colleges and universities. The very definite nature of this responsibility plus the greater selection which West Point has been able to exercise would suggest the possibility that cadet personnel problems would differ materially in number and in nature from the student personnel problems encountered in the customary college or university. It is the purpose of this paper to point out, only in a general way at this time, similarities and contrasts in the area of individual adjustment problems and to describe briefly certain procedures related to these problems.

Excluded from major consideration at this time are the matters of records and the highly significant area of cadet selection. With respect to the latter, suffice it to say that those who actually enter West Point represent between one-fourth and one-third of those who reach the stage of taking the qualifying examinations. It should be pointed out further that certain of those candidates who take such examinations already have gone through the preliminary screening examinations conducted by their Congressmen or by the commanding generals of service or foreign commands.

Traditionally, men have been selected for West Point primarily in terms of their academic ability or aptitude and have

---

<sup>1</sup>The opinions and assertions in this paper are those of the writer and are not to be construed as official or reflecting the views of the United States Military Academy or of the War Department.

taken their commissions upon graduation primarily in an order determined within each class essentially by the academic competence they manifested while at the Academy. While it is true that the men had to pass medical examinations for entrance, the criteria and methods employed have been such as to necessitate the addition of tests of physical efficiency or of physical aptitude in order to obtain more significant screening. The class entering next fall, for instance, will be the first class in the history of West Point to have passed through such definitive physical screening.

Contrary to the expectations of the uninformed, a group of cadets which has been selected in terms of the customary evidence of academic ability and aptitude and screened by a medical examination considerably more exacting than is the case in most colleges or universities, still is found to have in it men who are not emotionally suited to the process of learning to be military leaders. Inaugurated this spring for use on the candidates for entrance next fall was a biographical inventory in which was included a large number of items from the *Minnesota Multiphasic Personality Inventory*. The responses on this biographical inventory will be collected for at least two successive years, without being used for selective purposes, and will be studied for their discriminative value with respect to the emotional aspects of adjustment. On the basis of this and related studies on emotionality it is hoped to refine further the process of selecting future Army officers.

As yet no systematic and objective study has been made of the kinds, frequencies and degrees of cadet adjustment problems. Cadets who are dropped from the rolls of the Academy are academically deficient, or are physically deficient, or are lacking in leadership aptitude for the service. These inadequacies may be present either singly or in combination. In addition to these, some men drop out by resigning from the service. Included primarily among those who are dropped for inadequate aptitude for the service and among those who resign are many of the customary types of college maladjustment cases in which emotionality figures largely, and there is the usual overlapping of such emotional maladjustments with academic and physical inadequacies.



The conception of cadet personnel problems necessarily includes much more than the mortality aspects of the situation. Among the much larger number of men who remain there are those whose effectiveness is impaired as a result of unfortunate educational backgrounds, physical and/or emotional immaturity, or hampering personality traits. Viewed in terms of the entire Corps, the size of the group of cadets having problems of these or related types is probably relatively small. Still, they are among those who have been accepted for training on the assumption that they will profit from it to such an extent that they will become satisfactory Army leaders.

The following are presented merely as illustrative of the kinds of adjustment problems that have been found among cadets, some of whom have dropped out of the picture and some of whom may drop out because of their inability to profit from the help being given them. The order in which they appear bears no significance.

1. The typical bewildered adolescent, reasonably mature socially, but seeking a meaningful pattern of ideals, philosophy and religion. No moral question as such is apparent.

2. The son of a service man usually falling into either of these categories:

- a) The boy who is expected by his father, mother, uncle, or aunt, to carry on the military tradition of the family but who, for one reason or another, strongly prefers a civilian future.

- b) The boy who, having attended from four to nine different elementary schools and from two to five different secondary schools and having undergone considerable educational mutilation in the process, is interested in a military career but has serious difficulty with the academic phase of the program.

3. The son of psychologically inadequate parents who seek to compensate for their own shortcomings, either real or imaginary, by sending their immature offspring to become an Army officer, even though he may not be particularly interested in a military career.

4. The boy who seemingly strongly desires to become an Army officer but who has difficulty adjusting to the Academy program because:

- a) He breaks into tears when corrected.
- b) He is so apathetic, careless, or sluggish that he can not get along with his fellow cadets.
- c) He is so immature or overcompensatory in his behavior that his leadership ability is practically non-existent. Overlappings of these kinds of problems are, of course, quite common.

Let it be clearly understood that it is extremely probable that such maladjustments are found in only a small percentage of the cadet corps. Further, many of the cadets whose effectiveness is impaired by such problems are being helped materially and will develop into competent officers. Increasing recognition is being given to the existence of such problems and the thinking and action of more and more officers and cadets are being directed to the facilitation of the adjustment of cadets with such problems. In the not too distant future some of the more seriously handicapped candidates will be identified at the time of their seeking to enter and will be refused admittance—both for their own good and for the betterment of the Academy program.

It is the writer's thesis (1) that cadet personnel problems exist just as do such problems among college and university students, although the extent to which they exist is not now objectively known, and (2) that, ignoring the academic aspects of the Academy program at this time, the functional structure of the Corps of Cadets and the training problems inherent in such a situation make for a more effective handling of such adjustment problems than is true in most colleges and universities. The balance of this presentation will be devoted to the development of the second statement.

The training of cadets at West Point is provided by two Departments—the Academic and the Tactical. For tactical instruction, the Corps of Cadets is divided into 24 companies, each of from 85 to 95 men. Each of these companies is under the direct supervision of a carefully selected tactical officer who has had successful combat experience. Cadet officers are utilized in the operation of each of the 24 companies, as well as in the functioning of the two regiments and of the brigade. Within

the limits necessary for the proper functioning of the brigade, the 24 tactical officers are permitted, or even encouraged, to introduce and use such personnel methods as they regard as necessary or desirable. It is in the nature of this organization and its operation that certain of West Point's personnel advantages lie.

The homogeneity of vocational interests is obviously one factor contributing to more effective work with the cadets. Further, the ratio of one tactical officer to 85 to 95 cadets is superior to ratios existing between the average college and university faculty members and their instructional loads. But more significant than these is the fact that the tactical officers are more conscientiously concerned with and give more sustained personal attention to the problems of their cadets than does the average college or university faculty member. Brief descriptions of certain of the practices of these tactical officers will clarify this latter point. Some of these practices are common to all or most of the tactical officers; others have been introduced only recently by a few of the officers.

The officers tend systematically to interview their cadets. Their program consists, usually, of interviewing the fourth classmen (freshmen) very shortly after they first come into their companies, followed by the interviewing of the first classmen (seniors), since the cadet officers are in this group. Next come interviews with the second and third classmen, interspersed with interviews with any cadets who may be having trouble getting under way. These interviews, usually informal in nature unless a discipline situation necessitates otherwise, range from fifteen or twenty minutes to one and one-half hours in length. Cadets who are having some kind of difficulty are interviewed from once every other month to as many as three times a month. Not included in this description of interviewing is the extra personal attention which cadets get through extra instruction and coaching in academics. In some few cases, the cadets are referred to the station psychiatrist, and, since West Point now has a full-time professional psychologist for the first time in the history of any service academy, the tactical officers are with increasing frequency referring their adjustment cases to him.

The tactical officers use personal information on their cadets much more frequently than do college faculty members on their students. Varying somewhat from company to company, this information consists of the biographical data presented by the cadet at entrance, of autobiographies written by the cadets at the request of an officer, or of anecdotal reports varying in length from a sentence to a paragraph submitted, either periodically or spasmodically at the request of the tactical officer, on all or on certain cadets. In some companies a log book is kept on both the good and not-so-good accomplishments of the fourth classmen. In others, periodic performance reports are submitted either after cadets have performed assigned duties or when cadets have done something worth noting. As is usually true in such situations favorable reports tend not to be as frequent as do unfavorable reports, but it is pleasing to note that the proportion of favorable reports tends to be on the increase. In at least one company the cadets report on themselves, the officer of this company having cadets deficient (failing) in their academic subjects record their marks on a chart in his office until they become proficient.

Company cadet advisory boards on aptitude are used for the purpose of assisting the company tactical officer in identifying and working with those cadets who may be having adjustment difficulties. The manner in which such boards function varies among the companies. In some instances, they initiate action by reporting to the tactical officer on potential problem cadets and operate with only a modicum of officer supervision. In other companies, they await inquiries or suggestions for action from their tactical officers. These boards tend to regard their responsibility in a serious and objective manner, their basic conception being that they exist for the purpose of trying to understand and help maladjusted cadets rather than for the purpose of "jumping in for the kill."

In their extra-official relations, the tactical officers engage in practices which increase their effectiveness from the personnel standpoint. With a high degree of consistency, they invite all the cadets in at least one class, and in some instances all the cadets in their companies, to their quarters for social get-to-

gethers. To a much greater extent than is true of college faculty members, the officers visit their men in the post hospital. The officers take an active and sustained interest in the numerous intramural activities of their companies although the actual management of those activities rests in the hands of the cadets.

It will be recognized that there exists in a situation such as this a multitude of opportunities for the exercise of good personnel procedures. The activities just described as illustrative of the personnel activities of the tactical officers suggest certain of the ways in which those officers figure in the facilitation of cadet adjustment. The officers themselves are not only keenly aware of the direct part which they play in the personnel picture but also they are sensitive to the large part the utilization of cadets in these practices can play in training the cadets themselves to understand and to utilize desirable personnel procedures. In fact, this latter aspect is regarded as the more important by a number of the officers.

The statements that have been made thus far are not to be taken to mean either that the procedures are regarded as completely adequate or that the officers are covering all the aspects which they think should be covered. They are becoming increasingly aware, for instance, of the importance of some of the more elusive aspects of personality in the general problem of cadet maladjustment. They are keenly aware of the need for preventing the admittance of cadets who, because of obvious immaturity, of inadequate social adjustment, or of less obvious emotional problems, are likely to profit insufficiently from the training program of the Academy. The officers are recognizing increasingly the need for more and more reports of an anecdotal nature, especially those of a positive rather than of a negative nature.

One more aspect of the Corps' procedure is germane to this consideration, namely the manner of its dealing with the problem of any cadet's aptitude for the military service. To a limited extent, a cadet's final standing when he graduates is determined by this aptitude for the service. The proper determination of this aptitude is important not only for this reason but also for the purpose of identifying either those who may be

able to profit from special training or help in this area, or those cadets who are so lacking in this area that they do not appear likely ever to become acceptable Army officers. The careful thought, consideration and constructive efforts of the officers of the tactical department are coordinated on this problem of determining and striving to improve the cadet's aptitudes in the following manner.

The present system provides for each cadet's making judgments concerning every other cadet in his class and in his company with respect to their aptitude for service. Most recently the standing of each cadet within his company was arrived at by pooling the judgments of his classmates and the judgments of those men of other classes who were in a position to render meaningful judgments. On the basis of these rankings, combined in a number of cases with evidence of demerits, standing in a tactics course, and physical efficiency, the orders of aptitude within each class were determined.

Within each company the lowest ranking men in each class were studied to ascertain whether they were sufficiently inept to warrant further study. If they were so regarded, further evaluations of their behavior were made by having six cadets (three cadet officers and three other cadets who knew the cadets under study well enough to rate the subjects meaningfully) and the tactical officer fill out a cadet evaluation form which has been developed. In the light of the results of such ratings, plus the tactical officer's further judgment, the decision was reached as to the limitations or maladjustment of the cadet under study.

In case the cadet then appears seriously limited in promise, he is called before the Brigade Aptitude Board where the facts that have been obtained are presented to the cadet and he is given the opportunity to present his side of the case to the seven officers on the board. All available data are studied carefully to ascertain the kind and degree of the cadet's difficulty and a decision is reached as to whether the cadet should be recommended to the Commandant of Cadets for conditioning in aptitude, which necessitates a reconsideration of the cadet's case after he has had an opportunity to improve under guidance, for

discharge from the Academy, or for graduation without commissioning. On the basis of the recommendations of this board, the Commandant can then make similar recommendations to the Superintendent who in turn presents the matter to the Academic Board for final local action. Final War Department action must be taken on the recommendations of the Academic Board in cases of discharge or graduation without commissioning.

This rather lengthy description of the many steps through which the more serious cases of maladjustment must go has been presented for the purpose of showing the extensive consideration which is given to each serious case. All along the line, intensive study is made of each such cadet. The cadets and tactical officer in the cadet's company have seriously considered the cadet's strong and weak points and have taken steps to help the cadets. In many such cases, the cadets have been sent to the psychologist and/or the psychiatrist who will have at least studied the cadet. Many remedial suggestions are made by the psychologist to the cadet and to his tactical officer for the purpose of helping the cadet to respond more effectively to the training. The psychologist is also a member of the Brigade Aptitude Board and there goes further into the case with the cadet and with the members of the board.

It is apparent that there is a major difference between the learning problems of the college and university and those of the military phase of the Academy's training program, and that these present different possibilities from the standpoint of doing effective student and cadet personnel work. These differences contribute in large part to at least the possibility of a superior cadet personnel approach at the United States Military Academy.

# IMPORTANCE OF CRITERIA IN SELECTION AND TRAINING<sup>1</sup>

A. C. VAN DUSEN

Associate Professor of Psychology and Director, University College Bureau of  
Counseling, Northwestern University<sup>2</sup>

MILITARY psychologists participating in wartime personnel problems were frequently frustrated by the inadequacies of criteria available to validate their selection and training programs. Jenkins<sup>3</sup> has summarized the experiences psychologists had with such problems in Naval Aviation. He called attention to the factors which effect the reliability and validity of the criteria themselves which is in contrast to the customary concern for the reliability and validity of predictors. Likewise, inadequate criteria constitute a frequent problem in educational and industrial personnel practices. For example, in our consistent emphasis upon building better selection tests for professional schools, such as law, engineering, and medical schools, we tend to neglect the problem of how well grades in these schools (the usual criteria) reflect ultimate professional standing.

The positive steps taken toward the solution of typical problems encountered in Naval Aviation may suggest useful approaches to similar problems in civilian practice. This paper will examine grades on machine gun ranges as one criterion of aerial gunnery skill of Naval Aircrewmen. The analysis will show how such a criterion can be both unreliable and invalid and what positive steps can be taken to improve its usefulness.

This performance task has been chosen as an example in-

---

<sup>1</sup> The opinions expressed are those of the author and are not to be construed as official or reflecting the views of the Navy Department or the Naval service at large.

<sup>2</sup> The observations on this topic were made while the contributor was an Aviation Psychologist, Lt. USNR, attached to the Aviation Gunnery Training Standardization Unit.

<sup>3</sup> Jenkins, J. G., "Validity for What?" *Journal of Consulting Psychology*, X (1946), 93-98.



stead of a typical paper-pencil selection or achievement test, for two reasons. First, it is believed that many industrial skills could best be evaluated by performance tests if adequate and reliable standards could be devised. Secondly, just as in aerial gunnery, many industries prefer performance tasks as criteria of success in spite of their low reliability and validity. When such situations exist, the personnel specialists must improve these standards as much as possible. Therefore, attention will be called to the positive steps taken to improve the reliability and validity of this particular measurement of gunnery proficiency.

The task consisted of firing from a stationary ground position at a moving banner target which traveled over an oval course. Firing positions were about 1000 feet in front of the long side of the oval target track giving the gunner an opportunity to fire when the target was either on the near or the far side of the oval. It was assumed that this task was realistic enough to be considered a good criterion for skill as a gunner. To be sure, he used the actual equipment he would later use in the air. He fired at a moving target and received a very objective score in terms of hits in the target. In addition his instructor graded his ability to handle his gun and to remedy malfunctions when they interrupted his firing.

### *Conditions Effecting Criterion Reliability*

A close analysis, however, revealed that this criterion as it existed was inadequate as a measure of gunnery skill. A few of the conditions which lowered the reliability of the task will be cited here together with the positive steps taken to improve the reliability of this standard.

First, firing scores were unreliable because firing range conditions made stability in the number of hits scored impossible. Although several factors contributed to this instability only two will be referred to here. Slight differences in the condition and maintenance of guns account for wide variations in dispersion of bullet patterns such that one gunner's fire would be highly concentrated while the bullets of another would scatter over a very wide area. Variations in alignment of the gun sight

with the bore of the gun was additional cause for inconsistent scores. The positive action taken here was to bring to the maintenance department a better understanding of its part in the total training objectives and to encourage its replacement of worn gun parts and strict attention to mechanical adjustments. A department's success had too often been measured in terms of how long its gunnery equipment could survive under an increasing allotment of ammunition per gunner without regard to the precision of adjustment necessary for accurate firing.

For the type of gun range discussed here, 10 per cent hits was an above-average score. Tests of such high difficulty do not allow for discriminations in various levels of skill. It was found that the reliability could be raised by decreasing the difficulty of the task. This called for a modification of the physical situation, namely, an increase in the size of the target and a reduction in its distance from the gunner.

Even with the most favorable range conditions practicable, the failure to extend firing experience with any one kind of gun mount beyond the initial stages of learning was also responsible for the low reliability of firing scores. To alleviate this condition recommendations were made for an increased number of rounds to be fired by each man from each type of gun mount or turret. In addition, where it was possible, specialization of firing training was recommended in order to increase a man's training on the type of equipment he probably would use in his combat assignment.

As Jenkins points out, the recording of grades by personnel who do not observe the performance or who are inadequate to judge it, constitutes another factor responsible for the low reliability of a criterion. This point was also illustrated on our moving target range. Since several gunners fired simultaneously at one target, each man's ammunition was painted a different color so that each one's hits could be identified. To check on the accuracy of scoring, certain colors were systematically withheld, but scores for these colors continued to be reported.

An investigation revealed that the two factors largely re-

sponsible for this condition were the poor morale of the target scoring crews and the inadequate explanation to them of the importance of accuracy in scoring. The installation of improved incentives for this relatively boring task along with a periodic check on accuracy led to improved records.

To secure reliable measures in training was very difficult. However, because of the much greater difficulties in obtaining reliable combat criteria against which to check practices in training, it was necessary to make certain assumptions concerning the validity of the training content. One of these assumptions was that a gunner's performance with the guns he would later use in combat was very important in spite of any differences which might exist between ground firing and that in aerial combat. Following this assumption it was of prime importance to make firing scores as reliable as possible despite the unresolved questions about their validity.

#### *Conditions Affecting Criterion Validity*

As a parallel development to the preceding section, a few of the factors influencing the *validity* of machine gun range firing as a criterion of aerial gunnery skill will be cited. Although the reliability of criteria used in training was of importance for the progress of the gunners, attention was constantly given to bringing the training requirements to resemble more closely combat tasks. A brief reference will be made to the personnel specialists' attempts to improve the validity of this standard.

Perhaps the most significant factor which tended to make range firing an invalid criterion for gunnery skill was its lack of resemblance to the aerial firing task when a gunner's bomber was under attack. This lack of similarity was largely due to the insurmountable difficulty of simulating with a ground target the sighting deflections required in defending aerial attacks. Numerous modifications of target speed and path were attempted as partial solutions to this problem. Air firing at towed sleeves for the most part came no closer to representing the realistic combat sighting problems than did ground ranges, although such practice did have the advantage of being

airborne. Motion pictures of fighter attacks provided realistic sighting deflections but did not have the airborne advantage. As an outgrowth of numerous attempts to combine realistic sighting tasks with gun handling in the air, special camera guns were developed and perfected as a practical achievement test for the knowledge of sighting theory under airborne simulated fighter attacks.

To be an efficient gunner was, however, only one responsibility of the Naval aircrewman. Reports from combat aircrewmen revealed that most of their time was spent in duties related to their rating as radiomen or mechanics and to all-round aircrew jobs instead of in gunnery. It might be argued by some that gunnery performance was, therefore, an invalid criterion, since quantitatively it constituted only a small part of the total task required of the aircrewman. A qualitative evaluation of the consequences of his inability quickly and correctly to apply his skill in the few instances when he was being attacked immediately suggests that qualitative consideration may be even more important than the quantitative ones in establishing the validity of a given criterion.

Range firing, like most other aspects of gunnery training, was continually being modified by the introduction of different gun mounts and of equipment as new planes and their gun installations were perfected. The use of obsolete gunnery equipment for training might invalidate this performance as a predictor of success on newer and different equipment. To guard against invalidating this criterion by having training out of phase with fleet requirements constant liaison was maintained with agencies directing equipment modification for combat use.

### *Summary*

The examples of factors which influenced the reliability and validity of a single standard of gunnery success in Naval aircrew training serve to remind us of the importance of examining systematically the adequacy of the criterion itself in any selection or training program. These particular factors have been mentioned because they are frequently operative in criteria used in educational and industrial situations. A list of them

may help in the recognition of inadequacies of criteria in our civilian problems. The positive steps taken to improve the Naval criterion may suggest methods for improving the quality of civilian criteria when they are similarly deficient.

When the *reliability* of a criterion is low, consider the feasibility of the conditions listed below and the recommended action.

1. Conditions inherent in the task itself.—Determine the nature of this difficulty, perhaps a modification in the specifications of the required task or in the equipment used in its measurement may be made. Encourage supervisory personnel to maintain the measuring instruments in order to permit consistent performance.

2. Insufficient sample of performance.—Where possible increase the length of the performance task.

3. Measurement in the initial stages of learning.—Where practical do not attempt evaluation until there has been adequate opportunity to learn.

4. Inadequate records or incompetent judgments.—Select and use the most competent judges available and provide detailed instructions and standard methods for keeping records. Provide incentives to follow the instructions and then systematically check to see that they are being followed.

When the *validity* of a criterion is low consider the feasibility that one of the conditions listed below is operative and the applicability of the suggested action.

1. Lack of similarity in achievement measured and job success.—Analyze the job to identify factors consistently important to job success and provide training and achievement measures related to these factors.

2. Measurement of only a small percentage of the factors related to the total job.—Make the measures as representative as possible of the factors important to the total job. A qualitative as well as a quantitative analysis of importance of each factor should be made.

3. Changes in job requirements.—Keep informed of changes occurring in job requirements and continually modify training

and achievement measures to keep them in phase with these changes.

It is evident that some of the low validity coefficients reported for predictors logically related to the criterion task may result as readily from inadequacies of the standard as from those of the predictor. Civilian counterparts to each of the criterion inadequacies found in this military situation may be identified once the personnel investigator is oriented to be as sensitive to the rigorous evaluation of the criterion, as he has traditionally become in evaluating its predictors.

# DEFECTS IN EDUCATIONAL AND PERSONNEL SYSTEMS AS REFLECTED IN MILITARY EXPERIENCES

WILBUR S GREGORY

Head, Counseling and Testing Services, University of New Mexico

WHEN invited to participate in this symposium on the contributions of military testing to educational and industrial personnel work, my first reaction was negative because I believed that the military testing programs in which I participated made little or no contribution to civilian personnel work.

There are several factors which support this negative attitude:

1. For the most part, military testing programs were organized and developed by persons who received their training and experience in test construction, personnel work, and clinical psychological procedures before the war. These persons applied their previously acquired knowledge and skills to the handling of military personnel problems. It is questionable that many new techniques or procedures were developed which can be applied to peacetime civilian personnel procedures or problems. In other words, most of the talks and published articles on contributions of military procedures to peacetime civilian procedures would appear to be described more accurately as summaries of contributions of civilian educational and industrial personnel procedures and techniques to military personnel problems, not vice versa. If the horse of techniques and principles developed before the war pulled the wagon-load of military personnel problems, let us give the horse his due credit and not accuse him of being pushed by the wagon.

2. Many military personnel problems are peculiar to the wartime situation, the highly specialized nature of many military occupations, the administrative problems and procedures

of the military organizations, the race against time and the enemy, and other aspects of the military situation. These peculiarities differ so much from civilian problems and activities that the possibility of "transfer" from military procedures to civilian procedures is definitely limited. For example, it is true that many new tests were developed during the war which proved to be valid and reliable in selecting personnel for specialized types of military duty. However, many of these tests are of use only for selecting persons for specialized military duties and have no direct use or place in civilian educational or industrial personnel work. For example, what peacetime use is there for tests developed to select gunners, bombardiers, or navigators?

Of necessity, military testing programs were limited to military tasks, and only in those instances in which the military tasks are actually performed in peacetime occupations or training procedures can the tests be used in civilian personnel work. This principle of "transfer" of techniques applies to other phases of personnel work in addition to the development of specific tests.

3. Although I and many other professional personnel workers and psychologists may evaluate our military service as "useful experience," it is at least debatable that the military experience was more useful or valuable than the equivalent number of years of peacetime experience would have been. This statement should not be interpreted as a complaint. Most of us went into the service (and will do so again, if necessary) because we hoped we could contribute something to the war effort, not because we wanted to profit from it.

However, the negative attitude outlined above regarding the contributions of military experience to educational and industrial personnel work is not a complete view of the question. In the first place, there were developed during the war some techniques which can be used in civilian personnel work and research. Secondly, problems which arose in handling the personnel and training problems in the services brought sharply to our attention some of the defects and weaknesses in our educational and personnel systems.



These weaknesses and defects are not "new discoveries." A review of the educational, personnel, and psychological publications of the past thirty years yields discussions of all of them. However, they are weaknesses or defects with which we have not dealt effectively to date in our educational and personnel systems. The following discussion of several of these problems endeavors to illustrate how they arose as problems in a minor phase of A. A. F. training and personnel work, and to present the need for research in similar areas in order to improve our civilian educational and personnel systems.

1. Grading on the basis of a distribution curve and failing a small percentage of persons at the lower end of the distribution is common policy in our school systems. When grading procedures were set up in the aerial gunnery schools in the Army Air Forces, the question of "grading on a curve" was raised. How could anyone be certain that all of those who were graduated from gunnery schools were adequately skilled to be members of a combat bomber crew? If a gunner was not adequately skilled, the lives of all the men in his crew were jeopardized. The "stakes" were obviously too high to justify grading on a curve and blindly assuming that those above a certain percentile rank in gunnery school averages could be classed as qualified for air crew membership.

The vital question was: How proficient must a gunner be in order to be classed as a qualified member of an air crew? In order to answer this question, men were sent overseas to secure data on the basis of which proficiency criteria could be determined. On the basis of combat proficiency requirements and other criteria, minimum proficiency requirements were set which were based on how skilled a gunner *should* be. The question of where he ranked on a percentile curve became irrelevant.

In peacetime education, both cultural and occupational, the lives of men are not at stake in evaluating the proficiency of persons in various courses or schools. However, is it not time that we began to determine how proficient a person should be in various skills rather than to give students passing grades merely because they rank above the lowest 5 per cent of the

group who are in the course of study they are taking? Is there not a need for business and industry to set minimum proficiency requirements which can be adopted by schools as criteria for passing or failing students? Is it not time for counselors and personnel workers to begin to conduct research regarding minimum proficiency requirements rather than to depend on distribution curves and percentile ranks?

2. Another interesting and practical psychological problem which arose in one of the training programs with which I was associated for several months was that of how much practice is necessary to be certain that a man really knows the skill or information he is attempting to acquire and to be certain that he will retain it for a specified period of time? This question arose when it was discovered that some of the training procedures used very early in the war were not adequate to insure that the men would retain skills they had learned in training in this country until they were overseas. The period which elapsed between the completion of training in this country and the participation in combat duty overseas was, in many instances, of several months' duration. As a result, studies were undertaken to determine the minimum amount of repetition of lectures, tests, practice sessions, and other procedures necessary in order to have men learn certain information and skills so well that they would remember them for a specified period of time.

I doubt that these studies were ever satisfactorily completed. Nevertheless, the implication for our civilian educational and personnel systems should be clear. Educators certainly know that one lecture on a topic or one practice session in the performance of a function does not guarantee the retention of the knowledge or skill for a very long period of time, even though they often ignore that fact in organizing college, high-school, and grade-school courses. Is it not time for us to start research on this type of question in our college and public-school teaching, instead of going ahead on the assumption that one lecture on a topic accomplishes much of lasting value, and is there not a need in personnel work for extensive research into the rate of forgetting and the reliability of memory rather than merely the reliability of the achievement or proficiency tests used?

3. A third problem facing both our educational and our personnel systems made a strong impression on me during the war. This problem is concerned with the lack of educational techniques and methods for developing character and social values in men and women and the lack of adequate personnel techniques for testing or diagnosing character and social values.

Morale problems arose at times because, as one officer stated, "Too many men do not feel a personal responsibility or obligation to do their part in the war effort." The majority of officers who cited this problem defined it as a lack of character in that it involved an unwillingness to accept social or moral obligations and responsibilities as a member of the national society.

In my work as a clinical psychologist with combat casualties, I was strongly impressed with the fact that a majority of the "combat fatigue" and psychoneurotic cases with whom I worked were men who were self-centered, selfish, and lacking in this feeling of obligation to do their part in the war. As a result, combat duty or any other duty which involved personal sacrifice resulted in a severe emotional conflict. This factor was not involved in all combat fatigue cases, but it is my impression that it was in many.

What methods have we developed for educating people to traits of character and social responsibility? Can we assume that our present educational system develops unselfish social responsibility and a balanced sense of human values? One trivial example of an incident that happened during the war may help to illustrate and to emphasize the negative answer to these questions. Most of us recognize that Christmas away from home in time of war is not the happiest of occasions for young people who usually are home for that occasion. Most personnel workers and statisticians agree, too, that the third number to the right of the decimal in a coefficient of correlation is not of vital importance. Yet an officer who had his Ph.D. degree refused to secure a Christmas leave for a young enlisted man because he wanted that enlisted man to recheck a series of coefficients to be sure that the third decimal place had been computed correctly in spite of the fact that the first two decimal places had already been rechecked!

That example is not used to prove a point but simply to raise and illustrate the question: Does our present educational system, in spite of its perfectionistic goals in certain areas, really accomplish much in the development of character and of human and social values? Do we know how to develop character and social responsibility?

There are many occupations in which character, social responsibility, and sensitivity to the happiness of others are as important as abilities, interests, skills, and personality traits. Yet we do not have any tests which attempt to evaluate or measure these tendencies, with the possible exception of the *Allport-Vernon Study of Values*.

The war experiences of many persons emphasized the need for the development of educational and personnel procedures for developing and evaluating character.

4. The fourth and last problem which I chose to mention in this brief paper involves the development and testing of personality tendencies.

Personnel workers are constantly faced with the matter of the important role that personality tendencies and motivation play in job and school adjustments. The importance of these factors was reemphasized by those who did personnel work or research in the services.

For example, one very crude study which I made of "job adjustments" of gunners resulted in the conclusion that the better gunners tended to be conscientious, serious-minded, mature men who took their jobs seriously and tried to improve their skills while on the job. The poor gunners were careless, immature, lacking in dependability. The poor gunners tended to be more confident of themselves (less conscious of or less willing to admit their deficiencies) than the better gunners. The better gunner appeared to be motivated by a love of flying, an antagonism for the enemy, by patriotism, and by other direct positive motives while the poorer gunner tended to be motivated by a desire for flight pay and prestige and by mere conformity to orders.

Although clinical psychologists are developing methods for adjusting or developing personality tendencies such as those

used in work with individuals, findings such as those outlined above re-emphasize the need for research on methods of developing those personality tendencies through our educational system.

Likewise, because factors such as these are so important in job and school adjustments, the need for research in testing or other diagnostic procedures regarding such personality tendencies as these is equally necessary. What traits are most important in job and school adjustments? How can we test or evaluate them?

In connection with this general topic of personality tendencies, the problems of development of leadership skills and of testing for leadership skills appeared to me to be a major area to which our pre-war educational and personnel systems contributed little to the military procedures. And it is questionable whether any positive contributions to educational and personnel systems were developed or discovered during the war. How can leadership skills be developed? How can proficiency in leadership be evaluated objectively? How can potential leaders be selected for training or leadership? These problems, emphasized by the war, are important civilian, industrial, political, social, educational research and practical problems.

There are only four general problems which my military experiences impressed upon me as areas for extensive research. We often hear the statement that "War does not solve any problems." Perhaps, some persons who were more fortunate than I did find the solutions to research problems during their military service. My military experience raised many questions, such as the four I have outlined above, which require extensive research. Perhaps, if the war does emphasize and stimulate the need for research on problems such as these, it will make a great contribution to educational and industrial training and personnel work.

## WHAT THE VETERAN THINKS OF EDUCATION

HAROLD M. MORRIS

Graduate Student, New York University

THIS is the result of a survey made in the fall of 1946. Two hundred and sixty-three student veterans were interviewed in an effort to determine how they felt about the education they were receiving. This study can hardly be considered adequate. Certainly no generalizations can be made on the basis of it. While over 250 veterans might seem like a lot to interview, it still is not nearly enough for an adequate sampling upon which to base a scientific survey. Considering this, then, my purpose is not to make any sweeping generalizations. It is even difficult to point out indications. The most I can hope to accomplish today is to have some of you start to wonder, as I did, whether something is wrong with our educational system.

The survey was made in New York City, on the campuses of three universities. The 263 veterans who were interviewed were asked what they thought of the education they were then receiving. The replies were not flattering. Indeed, only three of them expressed unqualified approval. The collected comments might suggest that perhaps not everyone needs just what the colleges are now offering.

Two hundred and twenty-three of the 263 students remarked that the courses are not practical. This undoubtedly was the most frequent criticism. Here are a few of the representative remarks:

I'm wasting my time listening to a lot of worthless material. It is all theory, with no practical applications.

Much of the theory is so much nonsense. The required study doesn't meet my needs. Much of it is repetitious and boring. The only practical thing which I am getting from my college work is a degree. To a large extent I am wasting my time. I wish I could get some functional material. I can't find a job on theory.

I have been going to college only a short time, but so far I haven't received what I expected. I can't think of one practical fact that I have learned.

It stinks! There is nothing practical about it.

I wish my instructors would spend some time with me and tell me what good it is going to do me to learn all of this stuff.

The next most frequent criticism is one which is in sharp contrast to the claims of colleges and professors. Few topics have had more attention in recent educational literature than the importance of knowing students as individuals and of adapting instruction to individual needs. Yet, many of the veterans said that professors are not interested in the needs of students as individuals. For instance:

The school doesn't seem to give a damn what the student needs. There is no adequate evaluation by the school and instructors regarding the needs of the students—the veterans in particular.

Instructors are talking over my head. They don't seem to understand the needs of the group.

Instructors don't even make the effort to know me and my problems. During the summer, most of them haven't even office hours. They are too damned impersonal.

I went to see my German Prof. the other day and introduced myself. "Oh yes" he said, "I remember you, you are seat 42." What does the guy think I am, a piece of furniture?

Instructors seem to live in a world of their own, without considering their students. They should have a more personal contact with their instructor.

I surely do wish they would put a little more personal guidance into their teaching.

Another frequent criticism was about teaching method. Apparently some of the college students think that present methods are antiquated, dull and ineffective. Here are some typical comments:

It might be a good idea to give orientation talks to instructors occasionally, to refresh their memories as to effective methods of teaching and grading—not to mention objectives.

Much too much emphasis is given to reports.

I thought it was generally conceded that the lecture method of instruction was far inferior to many more progressive methods.

I feel like a woman at a nylon sale. We all seem to be struggling for marks and grades at the bargain counter of school.

Hell, I'm too busy writing reports and cramming for exams to learn anything functional.

The only thing some of my professors use the blackboard for is to lean against.

It is getting so I wouldn't miss some of my classes—I need the sleep.

Any creative thought on the part of students is discouraged. My papers which represent ideas and views which I sweated over and believe in are returned with caustic remarks, asking for proper documentation, footnotes and references.

There were only thirty or forty veterans who were able to express their thoughts adequately regarding the challenge of One World. Somehow most of them found it difficult to transpose their thoughts into words, yet the majority of them seemed to feel that education has not accepted the challenge. A few of them were surprisingly eloquent. They put into words what the others felt. They expressed themselves in words like these:

We seem to be educating for an egocentric society rather than for a world culture. It is just as feasible to have human dignity within a world culture as it is to have it in a regional community. It is the duty of the school to see to it that students recognize the fact that they are living in a world community.

I am a chaplain, and I have been one during the war. I am coming back to school to see how I can do a better job of guidance with the members of my congregation. I am greatly disappointed with the aims of education. It seems to me we must educate people to say "We are citizens of the world," with the same pride as they say "We are Americans" or "We come from California."

I suppose that school officials know best, but isn't it strange that in view of such profound world changes, colleges still hire such dyspeptic little men to discuss septic little details.

Not one of my classes ever attempted to discuss our place in a world culture. Even philosophy classes dwell on moth-eaten ideas of morals and ethics which hardly applied even when the world was a simpler place.

From the choice of subject matter at school, one might think that there was no such thing as war or an atomic bomb.

There were numerous criticisms about the G.I. Bill. Evidently it has not solved the veterans' problems entirely. Free tuition and books, plus \$65 a month, is a far more generous scholarship than most prewar students enjoyed unless they



played football. But take out room and board at today's prices and little is left for clothing, not to mention doctor bills and haircuts and toothpaste. Without exception every veteran interviewed who was receiving G.I. benefits, was getting supplementary funds from some other source. They were particularly bitter about delays in payments, restrictions upon their freedom to earn extra money in their spare time, and the housing situation.

Of course all of these criticisms cannot be taken at their face value. The student veteran sees only his own side of the picture. He has little opportunity to view the other side. A great deal of the invective might be due to a resentment over poor grades. Some of his criticisms are even contradictory, such as his desire to build a better world and his impatience with instruction that is not immediately applicable to his personal problems.

But one can make generous allowance for all qualifying considerations and still have left a significant and disturbingly widespread adverse criticism of college education.

What I have to say now is my personal opinion. I cannot possibly be sure that I am right. I am not even sure that anything is wrong though my sincerest personal conviction tells me that there *is* something wrong.

Actually I cannot help feeling that I should not be here now. I am not supposed to be telling you what to do. You have been telling me for the last ten years, through your books and classes. If I do seem a little hesitant, it is probably because now that I am realizing the college boy's dream of getting all of the professors together to "tell them off"—I don't know how to go about it.

What my opinion is worth I don't know. I am still only a student and far from sure of my ideas. My practical educational experience has been very limited. First of all, I am looking forward to becoming a teacher and perhaps when I do my ideas will change. Perhaps teaching will have a sobering effect upon me. But now I am still young and I see everything rather simply, as perhaps all young men do. We have our angry confident solutions to problems and whoever criticizes them seems blundering and blind. Perhaps it is so with these veterans also.

They are young too, and because some of them have for fleeting moments seen glimpses of truth, they felt justified in making sweeping indictments. To us youngsters, with the positiveness and the rationalization of youth, solutions come easy. Do this and that—and everything will be allright.

Perhaps that is so with me. I have my solution also, based partly upon what these men have told me, and partly upon my long years of experience as a student and my constant association with many other students. It would be all too easy to say: "Get out those decrepit men, who have grown old in discontent. Those men who ponder and plan to mend earth's sorrows by discussing yesterday with youth alive today, are perhaps too conservative for today's problems. Ambition, hope and love seem strangely weak and perishable things to them." The poet says, "I will believe in life while I am young, for once grown old there is no believing."

Yes, that would be easy but it would also be ineffective and impractical. I feel now, too, that perhaps I would sound more convincing if I were to use profound pedagogical jargon but that, too, is something which I believe we must get away from. Students today do not understand the cant of the profession. What they know is that they want to be doctors, engineers or social workers. They know that they want to earn enough money to buy an automobile and perhaps an airplane. They know that floating around just out of their understanding are tremendous ideas and concepts which they want to understand. They also know, even as you do, that they will never grasp these concepts by being told about them. It is far from adequate to have a man get up and talk about them. They must experience them for themselves.

Today in the schools of America the concept of guidance in education is becoming more and more widespread. I feel that this concept, carried to its fullest development is an answer to our problem. Perhaps not the complete answer, but certainly a step toward an effective solution. Testing, interviewing, and personality counseling are a part of it, but I feel that if a guidance program is to be most efficient, it must go even further. The average college student who goes to school for four years

has about sixty instructors. Each instructor has about four classes a day, and about forty students in each class. Surely with this contact each student and each instructor can get to know each other better if an effort is made.

It is this contact between instructor and student that is needed to accomplish the profound aims of education. In each person's life there can be only one or two experiences or persons which can have a most profound effect on him. In the student's life it should be the teacher. It does not require all of his instructors, but having a special room set aside with a label on the door marked "Advisors Office" is not enough. If each instructor can take but one or two students a term and get down on his level, then real guidance can take place. I have seen so many students (and I have had the same experience myself) suddenly find themselves lifted to high planes of new understanding and ambition by the efforts of an instructor. But even the large number that I know are but an infinitesimal number when compared to the whole of the student population. This contact between student and instructor in order to be effective must not be an impersonal affair. The psychological implications of having a professor write a letter to his student while on vacation can only be appreciated if one can peep into the mind of the student as he reads the letter. How many instructors have invited their students to their homes and how many have visited the homes of their students? Are professors too impressed with the dignity of their titles to go swimming with their students or to beat them in a game of bridge? The day that every student knows one member of the faculty to whom he may go for criticism and the day that every faculty member may call two or three of his students by a familiar name probably will be the day when the guidance movement will reach its peak of efficiency.

The personalization must not be an insipid thing, but a dynamic well-planned project, with objectives and methods all charted as carefully as lecture notes. Too few teachers have started in to examine in class the lives which their students are preparing to live. Too few have begun to teach values, human and social, not to mention love. Too many medieval-minded

professors view their students as abstractions, having no bodies, hopes, desires and loves. There have been many explanations and complicated cures for the ills of education. Little men with Van Dyke beards have poked every bone and every inch of educational skin. Then they have shaken their learned heads and prescribed pills for our immortal souls. All that time it is not the patients who have been sick—it is the doctors. They are the ones who need to be re-educated.

Without being in a particular situation, I cannot devise an administrative procedure which would assure every student of having a friend on the faculty. But considering the many other complicated administrative measures which have been initiated successfully by the policy-makers of a school, surely they can devise some way to insure the success of a program like this, if it is deemed worthwhile.

Of course there are many arguments that will be presented. I can hear the protests at faculty meetings. I can hear terms like "insufficient time," "family responsibility," "pay scales," etc., floating around the room. I may be wrong but I feel that this is the type of program that is needed vitally today in order to make really functional institutions out of our colleges and universities.

Besides, who knows, maybe even the professors will learn something.

# IMPRESSIONS OF STUDENT PERSONNEL WORK IN GERMAN UNIVERSITIES AND IMPLICATIONS FOR AMERICA<sup>1</sup>

E. G. WILLIAMSON

Dean of Students, University of Minnesota

Our small technical mission to Germany was organized by Dr. L. Dewey Anderson of the Department of Commerce to exploit industrial and technological developments in Germany during the Nazi regime and particularly during the war period. Our special problem was to study developments in industrial psychology and personnel work in industry, including vocational guidance, technical training in schools and universities, training on the job, assignment of man-power and many other phases of the program of utilization of man-power. The mission was composed of four psychologists, one economist, and one labor organizer. We were located in Höchst near Frankfurt but we traveled extensively in the British, French, and American Zones and, as a team, covered practically the whole of Germany except the Russian Zone.

Spending three months in Germany investigating any problem is not at all comparable to a sabbatical leave in America at the Huntington Library for investigation on a specific topic. German libraries are smashed and destroyed; people are scattered and difficult to locate as they move from one town to another without forwarding addresses. Investigations are further curtailed by broken-down or poorly repaired jeeps. Moreover, once individual German psychologists are found, they do not always readily discuss the details of their investigations. An American comes to appreciate the subtle results on human relations of the uniform of the conquering nation in dealing on

---

<sup>1</sup> The writer participated in this investigation of personnel work in Germany while on leave of absence from the University of Minnesota, September to December 1946.

either personal or professional basis with members of the conquered nation. Frequently, however, effective relationships among professional equals were established readily by means of the "courtesy cigarette" so well dramatized in newspaper accounts of the black-market operation. The courtesy cigarette has never been described as a basic technique in human relationships but to me it is as essential a rapport-establishing technique as are any of the ritual techniques of the non-directive school of counseling.

Occasionally one encounters a special kind of difficulty such as the one reported by Dr. Morris Viteles as he attempted to secure original research data validating psychological tests used in industrial firms in Germany. The most typically universal reply was that the validity data were no longer available for inspection because: (1) they were destroyed in the bombing by the British or American Airforces; or (2) the Russians liberated the test data to Moscow. One might expect that future historians who write about the development of aptitude tests and their uses in German industrial concerns during the war will make frequent reference to these two explanations for gaps in our knowledge about the validity of German tests.

One special problem of concern to me was an investigation of the uses of modern personnel methods in German universities in the selection, promotion, counseling and job placement of students, particularly those students in training for responsible technical positions in German industrial firms. My remarks tonight are derived largely from this part of my work for the mission.

### *The German Educational System*

A very brief review of some features of the German educational system will serve as an orientation for the remarks and generalizations I will make. Universal education in Germany is confined to the *Volkschule* or Elementary School. For the first four years, until the age of ten years, all children attend the same type of school. Then the vast majority go on to graduation from *Volkschule* at 14 or 15 years to be followed by part-time work and part-time vocational training. A small number graduate from *Mittelschule* and enter such occupations

as social work or office work. Another small number enroll in one of the *Gymnasien* or one of the other types of *Höhere Schule*. Graduates of the *Höhere Schulen* who pass the *Arbitur* may enroll in a *Universität* or in a *Technische Hochschule*. Essentially both educational and vocational initial decisions, and not always at the individual's choice, are made at age 10 years or perhaps in some cases at 14 years. Incidentally the heavy tuition required of applicants to *gymnasien* often was a significant factor in these career decisions. This appears to be equally true at the university level. For example, in contrast to the substantial number of self-supporting students in American public universities, between 20 and 25 per cent of the students at Marburg University are partly self-supporting, probably a significant increase over pre-war conditions. Apparently fewer students from low-income families enroll in college preparatory courses and eventually go to the universities.

Education may continue until the chronological age of 16 or 18 but it is radically different from that of America in that, as I say, the four years of common education is followed by specialization in some skilled trade followed by part-time continuation education or work on the job. On the other hand, some students enroll in a *gymnasium*, which corresponds to the college or college preparatory part of our public schools with special similarities to our private college preparatory schools. There are many types of vocational schools and many types of *gymnasien*. I need not list all types of the vocational schools, but the different types of the *gymnasien*, or secondary schools, are important because of the separateness of the curricula. Instead of organizing a particular high-school with several curricula or departments, such as commercial, vocational and college preparatory, the Germans organize separate schools for different kinds of curricula and for different kinds of educational ideas and concepts. There are *gymnasien* which stress the classical languages and classical studies and there are those that specialize in and stress the modern languages and studies. There are *gymnasien* which give major attention to modern physical sciences and so on. But all of the *gymnasien* prepare the individual to pass the leaving examination, or *arbitur*, which

is required as evidence of having successfully studied the subject matters prescribed as necessary for study in the universities. The *abitur* taken at the close of the *gymnasia* is, in addition to the test of political purity, the single most important factor in the admissions requirements to German universities. This is a significant point with respect to the current situation in Germany since, during the war, many students failed to take the *abitur* or failed to finish the *gymnasia* studies before going to war. Without some relaxing of the rigid admission requirements, such veteran students now will not be able to enroll in universities. Admission by means of aptitude testing in lieu of high-school credits is not practiced in Germany. Indeed such an idea was not widely recognized as a possible solution of the current problem. Dr. Matz of Marburg was interested and asked for further information. But Dr. Grammel of Stuttgart *Technische Hochschule* felt very strongly that only by studying the carefully prescribed *gymnasia* subjects was a student prepared to study university subjects. Not aptitude but knowledge was Dr. Grammel's repeated formulation of a proper admission policy.

Within the German universities, specialization in the first year is the usual experience of students. A student enters directly upon his specialized studies and enrolls in a small number of general education courses or common requirements so characteristic of American universities and colleges. German universities are organized into faculties such as the Faculty of Law or the Faculty of Philosophy in which the usual liberal arts subjects in American universities are taught. In the University of Berlin there is a faculty of agriculture but for the most part the traditional faculties in the universities are law, medicine, philosophy and theology.

In the 26 German universities approximately 100,000 students were enrolled before the war but the Nazi gangsters reduced this to about 15,000. When one realizes that such a small per cent of the approximately 47,000,000 German population was enrolled in universities during any one year, then one realizes that very influential factors have been operating, and still are today, to restrict university training to those indi-



viduals who possess racial and political purity as well as financial and social background. Vocational guidance or any other phase of student personnel work is thus greatly hampered in performing its true functions as we understand them in the United States. In evaluating the German educational system, we need to remind ourselves of Conant's dictum about American education: "... to the extent that educational opportunity is determined by family status, education in the modern world makes for social stratification."<sup>2</sup>

### *Studentenwerk*

The German equivalent of a broad program of college personnel services is called *Studentenwerk*. Following World War I a very significant development arose adjacent to, but not as an organic part of German universities. Public corporations, with some directing but unofficial membership from the German university's administration and faculty, were organized to provide certain services needed by students but not available within the rather rigid structure and frame-work of universities. Parenthetically it seemed to be characteristic of the German development of higher education that social institutions inherited from earlier periods are not often significantly changed in structure but are rather added to from time to time as new needs arise. The 19th century German concept of assigning to the student responsibility for his own out-of-class situations had persisted into the present century. Paulson summed up this attitude toward students and the policy of self responsibility in the words: "We must risk boys if we are to gain men."<sup>3</sup> But after the first World War, the economic stress of students was so severe that the risk of losing, not gaining, men from this policy was too great to be continued. Something had to be done to assist the enterprise of the student in finding rooms, food and social life. Self-responsibility could not cope with the economic situation of that period and some organized program was necessary. This historical departure from the pat-

<sup>2</sup> Conant, James Bryant. "Public Education and the Structure of American Society." *Teachers College Record*, XLVII (1945), 145-194.

<sup>3</sup> Paulsen, F. *The German Universities*. New York: Macmillan and Co, 1897 pp. 208-209.

tern of the ancient university, as I say, grew up alongside of, but not as an organic part of the university. When the Nazis came into power in 1933, they took over these public corporations, putting them under the Ministry of Internal Affairs. Today each university has its own local arrangements but perhaps the *Studentenwerk* at Marburg and the *Studentenhilfe* at Munster are fairly representative of these developments.

At Marburg the *Studentenwerk* provides the following services for students:

1. Housing, including helping the individual to find rooms in private homes in Marburg, especially those students who commute from long distances and also the refugee German student displaced from East Prussia or other provinces in the eastern part of Germany.
2. Health insurance for the improvement of health and the prevention of accidents, including inspection of housing, required physical examination of all students, hospital and medical services.
3. Dining rooms serving approximately 1800 meals per day. This is a very significant factor in the life of students because of the scarcity of food in Germany. You might be interested to know that for 65 *phennigs*, a dinner can be provided and for 55 *phennigs* a supper. There is a special meal for financially needy students costing 45 *phennigs*. When one realizes that there are 100 *phennigs* in one *mark* and that even at the over-valued Army exchange rate of 10 cents per *mark*, then one sees how little it costs to support German students according to American standards. A package of cigarettes is indeed a rich stipend in student life and the much photographed and still more sought after cigarette butt, while not a thing of beauty, is indeed a thing of life-value.
4. Dormitories are provided especially for the blind and physically handicapped students, most of whom are veterans. Incidentally 6 per cent of the students live in the dormitories and there are now 180 beds in the *Studentenwerk* dormitories.
5. Marburg has a *Studentenhaus* which is really the student's union providing physical facilities for social affairs. Obviously these buildings are by no means as pretentious as the lowliest of American college unions. But in the *Studentenhaus* there is an American library, a newspaper room and meeting rooms for faculty as well as a place for dances and parties.

6. Severely limited scholarships and free meals are available for financially needy students. Each student enrolling at Marburg pays 23 *marks* for each semester, 10 *marks* going for health benefits, 3 *marks* for the dormitory and building fund, 3 *marks* for special assessment for support of the blind students and for special readers. Part of this 23 *marks* provides scholarships and free meals for needy students, especially those dispossessed from the eastern provinces of Germany. Last semester (1946) 26,000 *Reich-mark* were given in scholarships and free meals and more than half this fund was raised by profits from the dining room, dormitories and social affairs. Approximately 10 percent of the student body of 3000 students were given scholarships by *Studentenwerk* and an additional 2 percent were given scholarships by the university or by townspeople.
7. The seventh service provided by *Studentenwerk* is an information service for new students which I shall discuss in a moment in another and broader context. Essentially this phase of *Studentenwerk* is an information service (*Akademische Berufsberatung*), or orientation to the requirements and courses offered by the university. In no German university could I discover that this advising was part of a counseling service or vocational guidance in the American sense of the word. Keller and Viteles describe such a program under the direction of Dr. Wienert in Berlin and by others in the universities in Darmstadt, Frankfurt, Halle, Hamburg, Cologne, Leipzig, Munich, and Tübingen. As I shall indicate later I thought at one time that this was the German equivalent of our counseling and testing bureaus but I found that this was not the case, since psychology in Germany has been applied to industrial personnel, but not to any appreciable extent within universities or high schools.

### *Counseling*

I turn now to a broader review of counseling in present-day Germany. In addition to the general information type of counseling referred to above, a second type is provided in the Federal Employment Offices located in the cities and large towns. No formally organized and systematic counseling or vocational guidance appears to take place in the schools, and as far as could be learned no professionally trained counselors (*Berufsberater*) are employed in the school system, at least not in the secondary schools. For the most part vocational guidance takes place in the last year of enrollment in the high

school. At this time the counselor from a nearby *Arbeitsamt* speaks to all the high-school seniors with respect to required registration in the employment office and the possibility of securing personal assistance from the *Berufsberater* in the *Arbeitsamt*. For the most part few if any of those students planning to enroll in universities come to the employment offices for vocational guidance and counseling. Other students come to secure continuing part-time training and work in the skilled trades or to enter apprentice training or to secure immediate employment. This was the same situation as reported by Keller and Viteles in 1937.<sup>4</sup> There appears to have been few changes made in this system during the Nazi regime and during the war.

Guidance at the time of leaving the high school is largely a job selection and assignment process within skilled trades. The schools do not participate in this counseling process except insofar as individual principals and teachers may advise a student or his parents. To a large extent the economic condition of the family still determines the choice of an occupational goal for the individual student. Fraulein M. Dreger and Dr. Merz of the *Arbeitsamt, Linneschule*, of Frankfurt reported that during the Nazi regime the best interests of the state rather than the best interests of the individual determined his occupational choice. This holds true in the present circumstances of national disintegration, which Dr. Stroux of Berlin University delicately called "The Catastrophe." Perhaps without clear recognition of the fact, counselors still participate in a process which subtly "assigns" some individuals to those occupations which have current shortages of personnel, have greatest social emergency needs which call for the least amount of financial and social background. It is to be hoped that this orientation and emphasis will be reversed when the present emergency is ended.

To repeat, vocational guidance in Germany is largely restricted to those entering the skilled trades. Many counselors employed in this guidance are not professionally trained counselors and do no testing or interpretation of aptitude tests.

---

<sup>4</sup>Keller, Franklin J. and Viteles, Morris S. *Vocational Guidance Throughout the World*. New York: W. W. Norton and Co., 1937. chapter IV.

Special testers, not all of whom are adequately trained in psychology, are attached to some of the employment offices and do the required testing. In this connection it was encouraging to talk with Dr. Gottschaldt of the University of Berlin and to learn of his current crusade to replace the present untrained testers with professionally trained psychologists. In this respect he is attempting to combine the occupational and psychological training of counselors much as we do in our American training programs for counselors. But this is one exception to the general rule in present-day Germany.

I may point out parenthetically that aptitude testing in Germany is not at all similar to that in America. The rigorous scientific and statistical validation processes so well known and established in America are not followed in Germany, with few exceptions. Tests are validated by simpler and less rigorous procedures. But that is another story well known to students of psychology.

With this inadequate and sketchy outline of some features of the context and social setting of personnel work in present-day Germany, let me turn to a discussion of implications. While not explicit in my discussions, I am assuming that you hear the incessant flat overtones and perceive the dull, depressing background of horrible physical destruction of beautiful buildings and the even more depressing destruction of the German people. These destructions constantly remind me of the Nazi gangsters who sought to elevate sadistic beastiality to the level of a governmental procedure of social control. Without such apperceptive background these remarks may seem to resemble the gibbering of an hysterical dean.

### *Some Implications*

A number of implications for American personnel workers occur to one who studies the program of man-power utilization in Germany. These are: (1) the failure to search aggressively for human talent; (2) the failure to provide extensive financial subsidy for talented individuals from lower economic classes; (3) the divorcement of counseling from education; (4) the inadequacy of education which neglects non-intellectual aspects

of student adjustment; and (5) the unexplored area of relationships between training for citizenship and personnel work. I shall discuss these implications as my own personal conclusions. Time and space do not permit me to buttress these conclusions with a myriad of personal observations. Neither does time permit either here nor did it permit in Germany an adequate determination of the representativeness of the observations on the basis of which these generalizations are made. I have never been more conscious of the theory of sampling with all that it implies than when I returned to America and started to generalize on the basis of what I observed in Germany. I am certain of what I saw, but whether what I saw is representative of what others saw and whether any of us observers are adequately sampling what actually occurred are questions which perplex one. But with these limitations in mind I shall discuss briefly these five implications.

*Failure to Search Aggressively for Talent.*—Despite the constitutional provisions of the Weimar Republic for scholarships ear-marked for talented individuals from low income families, relatively little seems to have been done to search aggressively for talent. Mr. Matz, Rektor of Marburg University, reported that there were rare cases of "poor" students being identified and encouraged by the lower schools or the *gymnasias* to go on to the university. An exception was sometimes found in the case of a Catholic clergyman who identified and encouraged a bright student from a poor family to study for the priesthood.

The Weimar Constitution of 1919 did contain provisions that aptitudes and inclinations of children should be the sole yardstick for determining the enrollment in secondary schools. But as Dr. Lorey of Frankfurt reported, "Only a limited number of financial facilities have been created in the way of scholarships and exemptions from tuition fees to enable highly talented children of parents of low-wage groups to enjoy the benefits of higher education."<sup>8</sup> Dr. Lorey goes on to say that there is not much evidence to be found of an aggressive systematic search for individuals possessing high aptitudes for scholastic or technical work.

<sup>8</sup> From a personal letter to the writer, December 8, 1946.

But there are no doubt numerous but unrecorded instances of teachers in high schools who single out talented young people and encourage them or even assist them in securing the means to attend the university. Such individual instances, however, do not substitute for a wide-spread and systematic search of talent from all economic levels of society. Moreover German aptitude testers have not been called upon for services in the development of state-wide or nation-wide testing programs conducted by universities or other societal agencies in an aggressive search for high-grade talent. The lack of such systematic efforts is all the more surprising in an advanced technological economy like Germany's in which a constant flow of highly trained man-power is necessary for the operation of that economy. In contrast with the democratic character of German universities of the 19th century,<sup>9</sup> resources for advanced technical training are now used for the professional training of those individuals whose financial resources and social backgrounds predispose them to elect and to become eligible for such advanced training. Such a system obviously fails to provide equal opportunity for many other individuals who probably possess intellectual aptitude equally high but who lack the financial and social qualifications for entrance first upon the *gymnasia* training and later upon the university training.

Thus it was that at the very time during the recent war that Germany was utilizing every known scientific device for the exploitation and utilization of its natural resources, and also at the same time utilizing modern and scientific methods of motivating adult workers to produce the necessities of war, that she continued to use ancient and time-honored methods of classifying, training and assigning individual students. This is, indeed, an amazing short-coming in the much-publicized program for utilization of man-power in a total war economy. It is equally a tragic commentary upon the academic insulation of technological psychologists who, apparently, saw no need or opportunity to apply personnel methods to the problems of selecting and training students in the university. I repeat that

---

<sup>9</sup> Paulsen, F. *The German Universities*. New York. Macmillan and Co., 1895. chapter III.

psychological methods were used in the selection and counseling of youths headed for the skilled trades, and for the selection of apprentices in the industrial concerns but not for selection of students in the *gymnasias* or universities.

Parenthetically it must be said that we in America should not point too stiff a finger of scorn at other nations. Speaking of America, Conant states that: "Anyone familiar with education knows that for a very considerable portion of the population it is the family financial status which places a ceiling on the educational ambitions of even the brilliant youth." It is of course true that we in America have an ever increasing number of city-wide and state-wide aptitude testing programs. It is also true that two national aptitude testing programs, the Edison talent testing program and the Pepsi Cola scholarship program, each year identify and subsidize many able students who, without such assistance, would be unable to secure the advanced training which they are capable of mastering. But in spite of such commendable efforts, education beyond high school is still closely related to the financial status of parents. Warner and his associates have demonstrated that among a thousand students all with IQ's of 117 or more, 95 per cent of those whose fathers' annual income was \$5000 or more planned to go to college. Of those whose fathers' annual income was \$1000 or less, only 25 per cent hoped to go to college.<sup>8</sup> A pre-war study revealed that in Minnesota, for each able student who continued formal education beyond high school, another student equally able terminated his education with his high school graduation.<sup>9</sup> Presumably limited financial resources was the chief reason. In Minnesota, and elsewhere, many bright students terminate their education at the close of high school while in Germany many able students conclude their training at the close of elementary or vocational schools. We in America

---

<sup>7</sup> Conant, *op. cit.*, p. 158.

<sup>8</sup> Warner, Lloyd and Associates. *Who Shall Be Educated?* New York: Harper and Bros., 1944.

<sup>9</sup> Anderson, G. Lester and Berning, T. J. "What Happens to High School Graduates?" *Studies in Higher Education*, Minneapolis: University of Minnesota, 1941. pp. 15-40.



have at least made this much gain in our new world in the conservation of superior talents through formal education<sup>10</sup>

*Limited Financial Subsidy of Human Resources.*—My second generalization is related to the first one. No aggressive, systematic and wide-spread search seems to be made by educators or personnel workers or any other public servant to find the necessary financial means to subsidize highly talented individuals from low economic groups. Again reference is frequently made to scholarships provided through the Weimar Republic for able but poor students. But it is reported that not more than 10 per cent of the students enrolled in German universities were on such scholarship provisions. Many thousands of equally able students were not able to go to the university because they could not pay the high tuition required to enter high school which is prerequisite to university training.

Dr. Walter Moede of Berlin reported in conference that during the Nazi regime he was called upon to test and screen the workers in factories in and around Berlin for the purpose of locating high-talented individuals who were to be subsidized by the state for university engineering training. Reference was made elsewhere with regard to other provisions for a small number of scholarships provided through *Studentenwerk*.

A hopeful outlook for the future is found in the report from Dr. A. L. Lorey of Frankfurt that the new constitution of the *Land Hessen* contains the following: (article 59)

Instruction will be free of charge in all public elementary, middle and secondary schools and in the universities. Equally free of charge will be the learning aide with the exception of those used at the *Hochschulen*. It will be provided by law that subsidies be paid to the talented children of socially under-privileged persons. It may be furthermore directed by law that appropriate tuition-fees be paid if the economic situation of the student, of his parents or of other persons responsible for supporting him warrants such payment. Admission to the middle and secondary schools will depend only on the aptitude of a student.

---

<sup>10</sup> Will French (chairman), *The Education of Youth in America*, Reprint from Teachers College Record, Vol. 48, No. 4, pp. 201-286, January, 1947. See especially Chapter IV, The Discovery of Outstanding Talent in Youth.

If financial appropriations are forthcoming, this constitutional provision may do much to give to Germany an adequate number of broadly and technically trained leaders whose social origins and attitude conditioning is sufficiently representative of all classes. The training of such leaders will facilitate the introduction of a greater degree of social mobility in the German economy.

*Counseling Divorced from Education.*—My third generalization is that counseling as described above actually reinforces the economic and social caste system in which those with wealth receive the highest amount of education, which in turn sets them poles apart from those who receive only the four years of elementary general education. Present-day counseling is pretty much restricted to the selection and assignment to skilled trades of students from lower economic groups. But even more significant is the fact that counseling in Germany is dominated by the social philosophy which places the welfare of the individual secondary to the needs of the economy or state. Counseling under these conditions can do little more than to add to the rigidity of the present social structure and caste system.

Indeed counseling may actually work against that social mobility which is so characteristic of America in which an individual with talent may rise, through education, above his social and economic origin. The basic social assumptions underlying American personnel work, at least in educational institutions, is that the counselor shall be free in his work from the bias and determining influence of economic, racial and religious intolerance and restrictions. In other words, the counselor shall be free to search for talent wherever he may find it and to enable the individual to use his talent in terms of his preferences and the amount and type of talent he possesses.

On the contrary, judging from my observations and readings, the basic social philosophy of German guidance places the needs and welfare of the state above those of the individual member. Obviously the use of the word "counseling" to cover these two different systems of personal assistance is inaccurate and misleading. The social purposes and the social results of counseling must be considered in evaluating as well as describ-

ing a system of counseling. And from this point of view we in America may need now and then to review the social consequences of our work. To too great an extent in America we may have occupied ourselves with technical analyses of the processes involved in counseling ignoring the social purposes and objectives to which these technical processes have been attached or for which they may have been prostituted by others. Perhaps we need to be aware of the danger that, as has sometimes happened with industrial personnel workers, educational personnel workers may be captured and exploited by those whose social philosophy does not permit maximum emphasis upon the welfare of the individual student. At the present time personnel work within some types of professional schools would seem to be restrictive in this direction.

Divorcements of professional and systematic counseling from education produces another serious deficiency in Germany which we in America have thus far avoided. Counseling thus divorced cannot become an integral part of the developmental sequence from childhood to adulthood in the broad sense of personality development. The emphasis must rather be upon bread earning in high school and routine catalogue-information giving in universities. Thus it is that counseling or guidance is reduced to a mere technician or clerical level of service and adds nothing to the force and direction of the individual's personal and social development. In these ways such a divorcement permits the German philosophy of "statism" to dominate over that of the importance of developing the individual in his own right. If I may assume the role of a mild-mannered Jeremiah for a moment, I would say that while we congratulate ourselves that we are more fortunate than our German colleagues, we should also seek ways of assuring ourselves that "It can't happen here." A generous and frequent dosage of social philosophy may help to correct our urge to narrow our "blindness" to technical and ritual aspects of our work.

*Inadequacy of the Intellectual Monastery Concept of Education.*—My fourth generalization is a commentary upon the inability of college professors to exclude from college life certain basic human needs and urges of the college student. Since the

invention of the modern form of the university "the professor's" major emphasis has been upon the intellectual development of the student as the main business of the university. In the Germany of the 19th century, intellectualism was associated with a theory and a method of developing a strong sense of self-responsibility, self-control and self-reliance. Paulsen describes it thus: "Freedom from outward compulsion is therefore the symbol of student days, the much vaunted academic freedom."<sup>11</sup>

This philosophy led to the emphasis upon giving to the individual student, in fact insisting, that he accept the full responsibility for his own life in all of its details outside of the classroom. In the classroom he was subject to instruction from an adult but outside the classroom he was entirely on his own responsibility. It should be pointed out, since it is sometimes forgotten by American advocates of the German philosophy of higher education, that during the early school years and even in the high school in Germany, totally the opposite philosophy dominated the relationship between the school and the individual pupil. This pre-university philosophy held that the student should be subjected to extremely strict discipline and regulation. But when he enrolled in the university, he was thrown entirely on his own resources usually without preparation. This procedure was said to produce a better type of adult who assumes full responsibility for his own life and actions.

In a very real sense, as Cowley has pointed out on numerous occasions, the 19th and the 20th century German universities disclaimed any responsibility for extra classroom needs, activities and adjustments of students. But important modifications of this extreme isolationist and over-emphasized intellectual point of view had begun even before World War I. Further extra-university developments were systematically organized after World War I, as was pointed out above in my discussion of *Studentenwerk*. Even in an intellectual monastery such as the German university was, or attempted to be, in the minds of the professors, certain non-intellectual or preparatory adjustments of the individual student had to be taken care of and

---

<sup>11</sup> Paulsen, F. *German Universities*. New York: Charles Scribner's Sons, 1906. pp. 265-66

could not be turned over to him as his own *sole* responsibility. But these developments remained organically and administratively peripheral to the university until the close of the recent war. This peripheral status obtained in American universities in the latter part of the 19th century and even in many institutions during this present century.

In 1919 economic conditions became so desperate among students in Frankfurt, Marburg, Berlin and numerous other universities, that official university attention was finally given to dwelling, food and social life. But with characteristic mental rigidity, German educators organized these programs apart from the universities, even though they were operated as services solely for the university student. It was as though the German universities tried to keep themselves historically pure by corporate divorcement of that agency which provided non-intellectual personnel services. But in these present confused and disruptive days following World War II, this historical purity concept seems to be lost sight of and now the universities, at least in the American and British Zones, are organically and legalistically incorporating these personnel services into the regular structures of the universities. What few opportunities I had to discuss this significant trend with the German professors led me to believe that they, like their American progeny, feel that this is a prostitution of the true university and the beginning of a lowering of academic standards and other types of departure from desirable institutional practices. But the students, with their age-old disregard for out-grown traditions and other types of ethnic vestiges, are most enthusiastic about the better roofs over their heads and the better food in their stomachs. This current instance of the truth that "the old order changes and gives way to the new," should be brought to the attention of our faculty colleagues who, once trained in German universities, have fixated in memory their own student days as unchanging in character even during Dr. Stroux's "catastrophe."

*Counseling and Citizenship.*—My fifth, and last, generalization describes a void, or unexplored area, of higher education in Germany. What I have to say could be said with equal

cogency of American universities. I refer to the unexplored area of relationships between counseling on the one hand and citizenship training on the other hand. There seems to be little formally organized provision in German universities for what we have come to call citizenship training. True, courses in History, Political Economy, Sociology, and other related topics are provided for the student who registers in a Faculty of Philosophy. But in a technical faculty, such as Medicine and Law and Engineering, ordinarily no such provisions are made for students. Such courses are, of course, provided for all students registered in the *gymnasias* and presumably they constitute his "general education" background. In the university he begins to specialize.

The training of highly proficient technical experts in engineering, physics, medicine and like, together with the professor's research, would seem to be the major purpose of a German university. Remnants of this same philosophy and point of view are found among the American professors who studied abroad in Germany during the latter part of the 19th century. As a result of this direct transmission of a concept of the function of higher education, in America we may have become so preoccupied with the training of technicians that we have forgotten that the college student must function as a citizen as well as a technical expert in his profession. This story has been told by so many Jeremiahs both before and during the recent Nazi gangster regime, that it needs no repetition here. Let me urge that you reread Paul Neureiter's 1934 analysis of the tragic failure of German universities to throw themselves into the then current economic, political and social problems and also about the resulting exploitation and discredit forced upon them by the Nazi and by their own university graduates. It is still shocking to recall that the burning of books was a contribution to Nazism by students of Berlin University.<sup>12</sup>

The implication I should like to point out is that the production of the best technical experts in Germany did not save her from or provide her with the protection from exploitation

---

<sup>12</sup> Neureiter, Paul R. "Hitlerism and the German Universities." *Journal of Higher Education*, V (1934), 264-270.

by Nazi gangsters. In fact one might collect evidence that technical experts produced by the university on the one hand and skilled but socially ignorant skilled tradesmen produced by the vocational schools on the other hand, left Germany with few politically sophisticated citizens. As a consequence politically and educationally naive persons accepted uncritically the myth of economic sufficiency, *lebensraum*, racial purity, Nordic superiority and other quackeries and nostrums advanced by the Nazis as arguments for giving them political and legal control of the nation. Technical proficiency does not seem to guarantee a sophisticated citizenry capable of selecting wise leaders who work for the benefit of all the people.

The significance of this generalization for personnel workers is self-evident. We have been largely preoccupied in America with the perfecting of our technical instruments and with the technical processes of counseling. It is possible that we have neglected one very significant part of our responsibility to our individual student. Indeed, we may have forgotten the student himself with our preoccupation with our technical problems, important though they are. After all, the student's major life adjustment actually takes place outside of the counseling locale. He is a citizen facing not merely emotional adjustment in marriage, and not merely demands upon his reading skills in the classroom. He is a citizen in the sense that he must do those things which will help to guarantee his continued freedom to exercise his talent and to permit other people to exercise their talents vocationally and in their own private and family lives. It becomes, therefore, an important problem for personnel workers to avoid a separation of personnel work from the training of the student as a self-governing citizen. Personnel workers should begin the careful exploration of the relationship between counseling on the one hand and preparation for self-governing citizenship on the other. Such an exploration leads immediately to the very cogent and real query of the contributions which personnel workers can make to the emerging field of general education. And I wish to close this account of my learning in Germany on that as-yet-unanswered query.

# GRADUATE TRAINING FOR EDUCATIONAL PERSONNEL WORK

CORINNE LA BARRE

Research Assistant, Western Personnel Institute

At the last national convention of the American College Personnel Association at San Francisco in 1942, Dr. E. G. Williamson, made this statement in his presidential address:

Most of us personnel workers were not specifically and professionally trained for our work other than having acquired generalized training in the supporting disciplines of education, psychology, sociology and mental hygiene. . . . The selection, induction and continuous in-service training of new personnel workers will continue to be a most difficult one as our profession increases in membership. We must recognize that one of the criteria of the development of a profession is the sources and methods used to replace workers as they retire. For the most part our college administrators now employ personnel workers trained in other fields, sometimes in fields almost entirely unrelated to the nature of personnel itself. Increasingly we are making available young and modernly trained persons but the source of supply is indeed small. Each one of us should assume some personal responsibility for increasing that supply of trained personnel workers.<sup>1</sup>

In the five years since, what has happened to this picture, re-touched largely from a wartime palette? For perspective, let us take a look at some of the original sketches of training for personnel work.

Until recent years, and to a great extent even today, personnel people have obtained their jobs in a hit and miss fashion, and not because of specialized training for this work. As a matter of fact, very little such training has been available. As in any new profession, training facilities become available only

---

<sup>1</sup>E. G. Williamson: "The Future Develops Out of the Past," *Report of the Nineteenth Annual Meeting of the American College Personnel Association*, San Francisco, 1942. pp. 4, 5.



after the demand for trained people is sufficient to warrant them.

Personnel work is still young, too young to have acquired set professional standards. These are still being created by the widely varying responsibilities and skills demanded in the employing agencies and by the experiments in methods and techniques still being carried on by the personnel workers themselves. Gradually these are being sifted down and a flexible but basic core of professional and training requirements is being established, as it has been established in other professions.

Training for personnel work in the past has consisted largely of getting experience on the job. Educational personnel work has been somewhat of an offshoot of other administrative or teaching positions. In many cases it has been a sort of extra-curricular activity added to the more clearly recognized responsibilities of faculty members. Consequently, while college training even on the graduate level has usually been a prerequisite for those engaged in personnel activities, it has been training directed toward instructional proficiency rather than toward personnel work.

In industry the practice of personnel-work training on the job has been especially prevalent. With the growth of large businesses, employment and employee relations duties have had to be turned over by top officials to other administrative officers. The need in these positions has been first of all for persons who knew the business—its jobs and its employees. These people were logically sought within the ranks of the company and this procedure still holds true for a large percentage of industrial personnel positions even today. In effect, then, practical experience has been the first consideration and academic training, though desirable, has been secondary.

Neither education nor industry, however, has been immune to scientific methods and to the development of basic techniques. Whereas the broad bases of psychology and sociology were once the areas of training offered nearest akin to personnel work, specialization has begun to set in. Gradually a body of knowledge is being built up which will best be assimilated

lated by future personnel workers through combined processes of schooling and experience.

Very little information is available concerning the early history of training for personnel work, either as to the programs of individual institutions or by compilation. Since personnel work came into existence as a profession after the first World War it is probably safe to assume that very little training was offered before then in other than some of these so-called "supporting disciplines." I made a survey in 1944 of institutions known to have organized training programs for personnel work at the graduate level, and found that Harvard University in 1916 was the first to offer training in educational personnel work. I understand since coming to this conference that there was even an earlier course started at Teachers College in 1913, for deans of women.

In the early twenties after the war the beginnings of other training programs became evident. In a study of training opportunities published by the Bureau of Vocational Information in 1924 a section devoted to personnel work lists five other colleges offering "organized programs of training for personnel work," the University of Michigan, the University of Minnesota, Bryn Mawr University, the University of Washington and the University of Wisconsin. These were all offering courses in industrial or governmental phases of personnel work.

While a number of other universities also offered single courses in personnel administration or industrial relations, and about twenty offered single courses in vocational guidance, this study goes on to point out that the emphasis on organized training was for industrial personnel work. The University of Chicago, Harvard University, the University of Michigan, Columbia University, and the University of Pennsylvania are the only institutions listed at that time as giving more than a single general course in vocational guidance. Of these, Columbia University was the only one offering a complete year's graduate work leading to a M.A. degree and a diploma in vocational guidance.

It should perhaps be made clear that the early concepts of educational personnel work for many were limited to vocational

guidance. It is, therefore, not surprising that much of the first training available was also limited to courses in Guidance or Vocational Guidance. While vocational guidance is an important phase of student personnel work, this phrase no longer adequately describes educational personnel work, especially at the college level. There is, however, still a tendency to refer to "guidance programs" or "vocational guidance" as constituting the personnel work being carried on in the high schools and junior colleges.

Since 1924 the opportunities for training in this field have greatly increased, with more colleges and universities offering educational personnel work training. Most institutions of higher education today have at least a few courses related to training for some or all areas of personnel work.

Field work in connection with course work in guidance and personnel was first offered at Teachers College, Columbia University, in the winter of 1927. The first course had two students, and in two more semesters the number had jumped to thirty. In 1928 a Vocational Guidance Clinic was also established at Teachers College offering practical work experience for advanced internes. Other early programs including in-service training at the graduate level were started at Syracuse University and at Stanford University, where "student deans" received training as residence directors and assistants in addition to courses leading to graduate degrees.

One of the first in-service training programs for personnel work not connected with an educational institution should also be mentioned. Under the sponsorship and direction of the Vocational Service for Juniors in New York City, an extensive program of field work experience was made available to a carefully chosen small group of applicants on a fellowship basis for the first time in 1930. Outside of a short seminar training period at the beginning and end, the program consisted of nine months of working in each of a number of different types of agencies known to be carrying on good personnel programs.

Of special interest in the Pacific Coast section of the country is a training program developed over a period of years by my own organization, the Western Personnel Institute in Pasadena.

This organization is a research center for colleges and universities in the eleven western states, and serves as a clearing house for information about educational personnel methods and occupations for college students. Two or three fellowships are given each year to graduates of colleges and universities holding membership in this organization and they become apprentices for a ten-month period at the headquarters office in Pasadena. Here they receive practical field experience and as much academic background as fills their individual needs in student personnel practices, plus a lesser degree of either public or industrial personnel methods in another agency. Work done in this program has been accepted as partial preparation for the Master's degree by Stanford University and by the Claremont Graduate School.

The increase in training for personnel work since the early twenties has been encouraging. More recently in the last decade, and especially since 1940, there has been a wave of new courses relating specifically to personnel work, and a rapid growth of graduate programs.

Wartime usage of this applied science of understanding, adjusting and promoting normal human relations by industry, government, education and the armed forces has swelled the ever-accumulating store of knowledge, techniques and tools now available to personnel workers. Besides, personnel work has become popular. The public has become familiar with such terms as "guidance," "testing" and "counseling"; it is learning to use these words and to ask for the services which they imply. You and I know, however, that a demand for services and the people to perform them does not always resolve itself into a demand for professionally trained personnel. Hence the interest of all of us here and especially those in the American College Personnel Association, in the training facilities that are presently available.

A sub-committee of the Committee on Student Personnel of the American Council on Education has been at work during the last year or so on a survey of the graduate training program available for those interested in educational personnel work. The results of this study will be published this year by the American Council on Education. I have been given permission to bring you a few highlights from this study.

A four-page questionnaire prepared by a sub-committee and approved by the Committee on Student Personnel Work of the American Council on Education was mailed to some 299 institutions offering graduate programs in this country. Returns were better than 90 per cent and it is believed that most institutions offering graduate training for personnel work responded.

The main areas covered in the questionnaire are: (1) Graduate Training Available, including the educational levels and special fields for which personnel training is available the degrees offered and the time required to complete them, and whether or not the preparation meets requirements for state counselors' certificates; (2) Departmental Majors; (3) Undergraduate and Experience Prerequisites; (4) Courses Offered; (5) Required Practice or In-service Training, including the kinds and whether it is available with students and/or other individuals, and the number of hours required; (6) Student Personnel Services in which training is available; (7) Faculty; and (8) Assistantships or Fellowships available, including number and amount of the stipend.

Like most questionnaires, the one used in this survey has many basic limitations. It does not determine much more than the bony structure of the training programs it attempts to analyze. Little indication is apparent as to the quality and quantity of meat covering the framework. In addition, this survey was hampered from the beginning by a lack of uniform terminology which is still apparent in the personnel work profession. Some may take exception to the terms used to describe certain phases of educational personnel training; as some found them difficult to use in answering the questionnaire.

Without wishing to try your patience on the statistical results of the survey, I am sure you will be interested to know that 110 colleges and universities answered "yes" to the question "Does your institution offer an *organized* program of graduate study for educational personnel work?" Nearly two-thirds of these schools are located east of the Mississippi; 20 are in the eleven western states; and the remaining 23 are in central states west of the Mississippi. New York has the

largest number of schools, 12. Texas and Ohio each have 8. Pennsylvania, Illinois and Massachusetts each have 7.

Fifty-nine of the 110 schools offered educational personnel work training and 17 additional colleges also offered graduate training for other kinds of personnel work including industrial, governmental, rehabilitation, etc. One hundred and forty-four offered no organized graduate work for personnel work of any kind, though a number indicated that they had graduate or undergraduate courses available.

In 7 states, Indiana, Missouri, New Jersey, New York, Ohio, Pennsylvania and Utah, some institutions indicated that their programs met the requirements for state counselors' credentials. I understand that there are 13 states in all and the territory of Puerto Rico which give counselors' credentials.

According to this survey considerable opportunity is available for in-service training along with academic work. Better than 80 per cent of those schools which give educational personnel work training offer practice experience, though very few indicate a prescribed number of practice hours required. All but 6 schools indicated some student personnel services with which practice training is available.

About one-third of the institutions indicated that they offered special scholarships or some form of financial assistance to those wishing to take graduate training in educational personnel work.

I think we would all agree that there is a need for adequate and effective training for the future personnel workers in education. Their selection in the first place is an important one, hinging on personal qualifications, academic background and work experience. The personal qualifications necessary to succeed in personnel work have not been defined professionally at this writing—and perhaps they never will be. It is certain, however, that all of those who are "fond of working with people" do not necessarily have the emotional stability and mature judgment required by personnel workers. A thoughtful approach to this aspect of selecting potential personnel workers is needed and research should be encouraged. Perhaps one of the personnel worker's own tools, psychological and personality

testing, may some day give further clues to the best selection procedures. I notice in a recent issue of *Occupations* a report on some experiments being made with a test designed to select potential vocational counselors. Undoubtedly there are others also.

Academic and experience prerequisites taken up in this survey do not indicate that any very great attempt has been made as yet to determine what should be the undergraduate preparation for future educational personnel workers. Whether work experience should be required before or during graduate training is a moot question, and needs further research.

A survey of this kind cannot show the quality or quantity content of the training programs offered by our colleges and universities. It cannot point up the lack of standardization which would be apparent between institutions upon a closer scrutiny of their training programs. While on the surface they appear to be similar some will offer more and better courses than others. Some graduate programs will be under the supervision of faculty members who have had many years of association with and interest in the student personnel field. Others will be but side-lines in institutions primarily interested in preparing students for other professions. The fact, however, that so many colleges and universities are now offering some kind of program for educational personnel work indicates the interest in the whole area of student personnel work. The next step would seem to be a visitation program by an organization such as ACPA to determine the extent and quality of the training programs.

Neither an increase in the number of training programs offered nor in the need for personnel workers will, however, insure that the work will be done by competently trained persons unless some measure of their proficiency is designed to meet minimum qualifications. This indicates the need at this time for some basic personnel-work standards, preferably on a national scale. The development of "counselors' certificates" by some states is a step in the right direction. Now is the time, however, for a professional student personnel organization, such

as the ACPA, to undertake the necessary research and study to establish minimum standards on a national scale. If it does not, it is almost certain that some evaluation will be made by people less qualified than professional workers in personnel.

Personnel work is one of the great professional potentials of a democracy. Its possibilities should be explored with imagination and courage and its foundations built soundly. The caliber of persons who choose this as their life work and the quality of training they receive are of primary importance to a profession which can contribute so much to society.



# FUTURE ENROLLMENTS IN U. S. INSTITUTIONS OF HIGHER EDUCATION

J. H. ROHRER

Associate Professor of Psychology and Research Assistant to the President,  
University of Oklahoma

THE reliable estimation of future enrollments is one of the major factors upon which depends the sagacious planning of future policy for institutions of higher learning. There are a number of approaches which can be made to this problem of estimation. Dean C. E. Partch<sup>1</sup> has provided an estimation which assumes a constant increment to the enrollment with each succeeding year. The resulting enrollment curve can be represented by a constant straight-line function. The validity of this estimate may be questioned on at least two points. The first question arises out of an examination of past enrollment data. If one plots enrollment in institutions of higher education for the years 1890 to 1940<sup>2</sup> the resulting smoothed curve drawn by inspection suggests a sigmoid or S-shaped function. That is, a curve obtained which is positively accelerated up to the 1920's, tends to flatten during the twenties and becomes negatively accelerated during the thirties. Such a curve suggests that the best assumption which could be made concerning future enrollment in U. S. institutions of higher education is that the increment to total enrollment will tend to decline in future years until a point is reached at which the enrollment is stabilized or perhaps decreases. It will be seen that the picture presented by the past enrollment is not in agreement with the Partch assumption of as constant increment.

---

<sup>1</sup> Partch, C. E. "Nation-Wide Estimate," *Journal of Higher Education*, XVI (1945), 241-246.

<sup>2</sup> These data have been summarized by the Statistical Division of the U. S. Office of Education. See: *Biennial Surveys of Education in the United States*, 1938-40 and 1941-42. Volume II, Chapter IV, "Statistics of Higher Education, 1939-40 and 1941-42." Washington: Government Printing Office, 1944.

A second question raised by Dean Partch's approach evolves from the logical extension of the assumption of a constant increment. If one projects the enrollment curve, on the assumption of a constant increment to each year's enrollment, a point will be reached at which the enrollment in U. S. institutions of higher education would be greater than the U. S. population itself! It is, of course, impractical to attempt to predict school enrollments over long periods of time because of the host of independent variables which may affect it in the future, but which are not known at present. For immediate predictive purposes the question raised of the logical extension of the assumption of a constant increment may be one of theoretical rather than of practical import.

TABLE 1  
*Estimation of Future Enrollments in U. S. Institutions of Higher Education  
as Obtained by the Curve-Fitting Technique*

Academic Year	Estimated Enrollment	Academic Year	Estimated Enrollment
1925-26	875,000	1953-54	1,842,000
1930-31	1,060,000	1954-55	1,850,000
1935-36	1,290,000	1955-56	1,857,000
1940-41	1,500,000	1956-57	1,864,000
1945-46	1,720,000	1957-58	1,871,000
1950-51	1,810,000	1958-59	1,877,500
1951-52	1,825,000	1959-60	1,884,000
1952-53	1,834,000	1960-61	1,890,000

A second approach which can be made to the problem of estimating future enrollments is one which involves a curve-fitting technique. The procedure consists in determining the functional mathematical relationship which exists between past enrollment and year,<sup>3</sup> and then in projecting this function to the year or years for which prediction is desired.

Table 1 presents data on future enrollments in institutions of higher education which were estimated by the curve-fitting procedure. The general equation for the curve is of the form,  $Y = ae^{bx}$ . This equation results in an exponential curve, sometimes called a "growth" curve because of the pervasiveness with which it is observed in growth data.

<sup>3</sup> For basic data see: U. S. Office of Education, *op. cit.*, p. 17.

From Table 1 it will be noted that the growth in enrollment is negatively accelerated and exhibits a marked tendency to flatten off around 1955. The validity of this trend is of considerable importance, because of its implications for future planning of physical plant installations, etc., in institutions of higher education. It is this negatively accelerated characteristic of the growth curve which leads to the marked discrepancy between the estimate presented herein and the estimate made by Dean Partch.<sup>4</sup> Partch estimates a total enrollment of 2,800,000 "by the time the veteran program is completed." If it is assumed that it will take until 1960 to complete veteran training, then the discrepancy between the two predictions for that year is 900,000 students. If the Partch estimate is correct it would indeed represent a phenomenal accomplishment on the part of higher education. An enrollment of 2,800,000 will be the equivalent of two out of every three persons in the 18 to 21 year age group attending college in 1960, as contrasted with a ratio of one out of seven in the same age group which were enrolled in institutions of higher education in 1940.

Evidence which indirectly supports the predicted decrease in enrollment growth made in this study may be obtained from an independent source—actuarial data on birth rates in the United States. The growth in total population of the United States is characterized by a negatively accelerated curve. However, certain age-group segments of the population are declining more rapidly in numbers than others. For example, in 1900 the age group of from zero to four years of age comprised approximately 12½ per cent of the total population; by 1940 the age group of from zero to four years of age comprised approximately 7½ per cent of the total population; by 1960 it is predicted that this age group will comprise only 5 per cent of the total population. The effect on college enrollment of this decline in the population will be felt some 18 to 20 years later and will be alleviated in part by the increased numbers of the population in the 18 to 21 year age group who will seek higher education. In 1900 4 per cent of the population in the 18 to 21 year age group attended institutions of higher education. By 1940 the

<sup>4</sup> Partch, *op. cit.*

percentage of the same age group had risen to 15 and 42/100ths per cent. If this growth trend continues until 1960, approximately 30 per cent of the U. S. population in the 18 to 21 year age group will be attending institutions of higher education. On the basis of these data, one can arrive at an independent estimate of the future enrollment in the institutions of higher education by taking the predicted increases in the 18-21 population which will attend college during a given year and multiplying that percentage by the actual number of the U. S. population which will be in the 18-21 year age group for the given year. This procedure results in essentially the same enrollment figures as those obtained by the curve-fitting procedure and hence may be interpreted as evidence which increases the level of confidence with which the first prediction can be accepted.

A third approach which can be made to the determination of the validity of the enrollment prediction presented in Table 1 grows out of an examination of data compiled by the Veterans Administration. Through the use of the predictive enrollment curve from which the data in Table 1 were obtained one can estimate the accumulated backlog of potential college students resulting from the war's interruption of potential college students.

This procedure involves cumulating the discrepancy between estimated and actual enrollments for the academic years 1940-42 to 1945-46 inclusively. The total estimated accumulation amounts to 2,590,868 potential students. For the academic year 1946-47, there were 360,000 more students enrolled in U. S. institutions of higher education than were predicted on the basis of "normal growth" enrollment.

The Veterans Administration, on Jan. 1, 1947, reported that there were approximately 2,050,000 veterans who had applied for educational benefits but who had deferred their educational plans. This 2,050,000 added to the 360,000 over-enrollment in 1946-47 totals 2,410,000—a figure in fairly close agreement with the two and one-half million estimated to have accumulated over the five-year war period.

Returning to Table 1, it will be seen that by 1950-51 when the influence of veteran enrollment should be on a marked

decline, it is estimated that institutions of higher education would, *due to normal growth*, have enrolled 300,000 more students than they had enrolled in the pre-war peak enrollment year of 1940. This is a 20 per cent increase over the 1940 enrollment. The influx of the returning veterans to the college campus has served to dramatize the need for expansion of facilities in higher education. The need itself has been developing over the past eighteen years as a result of the normal enrollment growth. It is extremely important to the future of higher education that the current nation-wide college and university emergency expansion program be accompanied by a permanent expansion program which will provide training and housing facilities for those increasing numbers of American youth who will be attending colleges in the years to come.

#### *Application of the Enrollment Prediction*

An application which may be made of the U. S. enrollment estimate is in estimating future enrollments of individual institutions. Despite the increase in the error of estimate which will result from such an application of the data on national enrollment, the estimate may serve a useful function for individual institutions by providing a base norm on which individual modifications, arising from conditions unique to the local situation, can be made. A technique for calculating this local norm is as follows: Determine by years the per cent contribution which the local institution has made to the national enrollment in institutions of higher education; determine the average of those percentages—this is your *norm percentage*; multiply this percentage by the estimated national enrollment for the given year (this data can be read from Table 1); the number obtained is the *norm enrollment* for the local institution for the given year. It is to be recognized that this *norm enrollment* is at best a rough approximation. Certain obviously important factors have not been taken into consideration: school leadership which may result in a more vigorous growth of the institution; the rise and/or decline of outstanding instructional departments; temporary economic conditions; fixed, stable enrollment in certain institutions of higher education, etc. The only justified

Unit (N.C.E.R.T.)

2735

tion for mentioning the procedure is a pragmatic one: the estimate of enrollment which is thus calculated is about the "best guess" which can be made in an area in which a "guess" is imperative if any planned institutional expansion program is to be embarked upon.

In closing, I would like to point out one major problem suggested by a consideration of the current and anticipated increased enrollment in higher education. This problem has to do with the body of knowledge which future freshmen populations will have when they enter college. The current trend of reduced college preparatory training at the secondary-school level, coupled with the increased number of high-school graduates who will attend college suggests that an increasingly large number of students will be attending our colleges and universities in future years who are deficient in basic academic skills, e.g., mathematics, communicative and language skills. There are at least three possible ways of alleviating the seriousness of this problem. These ways are: (1) Institutions of higher education can take over the task of training students in elementary basic skills—a function formerly assumed to be served by the secondary school; (2) A readjustment of college curricula could be made so that a freshman student would not be penalized by his deficiencies in fundamental academic skills; (3) A reorientation of the dominant current philosophy held in secondary education could take place so that they would once again assume the responsibility of developing fundamental, basic academic skills necessary for success in higher education.

## THE COLLEGE COMMUNITY AND OCCUPATIONAL INTEGRATION OF MINORITIES

FRANK S. LOESCHER

Placement Service, American Friends Service Committee

DURING the past year and a half I have been interviewing the heads of Philadelphia banks, insurance companies, department stores, industries, hospitals, colleges and schools to put before them the vital importance of equal opportunity in employment and upgrading. The findings of this race-relations project of the American Friends Service Committee may be helpful to college personnel people interested in the growing points of our democratic society.

But first a word concerning our reasons for establishing a non-fee service for minorities in Philadelphia, involving a placement service for Negro men and women and a counseling service to employers on techniques of implementing the principle of employment on merit, irrespective of race, creed or nationality background.

Placement work is not foreign to the American Friends Service Committee. For the past eight years we have been placing European refugees. During the war we assisted Americans of Japanese descent who were forced to leave their homes on the West Coast. Our work with Negro men and women, however, represents not so much a response to an emergency as a sensitizing of our consciences to a condition which has existed for generations. World War II awakened many Quakers to the gap between our testimonies and our practices.

Gunnar Myrdal, Pearl Buck, Lillian Smith and other writers have helped us to understand the heart of the problem of racial and religious minorities in the United States—segregation. Jews, Japanese, Chinese, Mexicans, Negroes, Catholics—the list is a long one—are *in* America but are not always a participating part of America.

I am in this work because I happened to read that extraordinary work, *An American Dilemma*, while spending a sabbatical at Fisk University where I learned first-hand the meaning of segregation and its concomitant, discrimination. I asked myself why are not educational, civic and religious groups helping to place trained Negro men and women not only for the sake of the colored students but even more for the sake of white men and women who have not had the opportunity to work with and to get to know Negroes of similar backgrounds and interests.

### *Employment Trends in the North*

The Philadelphia story is not headline material. Employers, with only one exception, have discussed sympathetically the employment and upgrading of Negro men and women, but few have changed their policies. Quaker and non-Quaker are reluctant to take a stand and to get their executive staff to carry out a policy of hiring and promotion on the basis of merit alone. Some say that they fear an unfavorable customer reaction; more say that they fear an unfavorable employee reaction. Nevertheless, doors are opening, notably in the department store field, in chain stores, in several industries, in hospitals, colleges, schools and libraries.<sup>1</sup>

Philadelphia lags in comparison with New York City where fair employment practices are becoming established patterns. The Chase National Bank, the Metropolitan Life Insurance Company, Macy's—to cite three institutions I visited recently—employ Negro men and women in a variety of positions. Before the passage of the anti-discrimination law Chase had a Negro woman in its library and Negro men and women in clerical positions. Macy's today has 25 colored sales people and numbers of clerical workers. Metropolitan did not begin until after the passage of the anti-discrimination law, but today they have Negro white collar workers in almost every occupation and in almost every department. There is no segregation. And there are no "problems."

<sup>1</sup> For the philosophy of this race relations project, its aims and methods, together with an analysis of the Negro employment situation in Philadelphia, see Loescher, Frank S., "The Placement Service of the American Friends Service Committee: A Technique in Race Relations." *Occupations*, XXV (1946), 90-93.



Throughout the country the movement of members of minority groups into the main stream of American life is accelerating. It is significant that more than 50 Negro men and women have taught on the faculties of Northern colleges and universities during the past two years.<sup>2</sup>

One of the incongruities of this whole situation is that along with the widespread pattern of discrimination there is a shortage of qualified Negroes for certain occupations, particularly clerical and technical.

Why are there not qualified people for these clerical and technical positions? First of all, willingness to consider Negroes is a recent development related to the increasing awareness and concern among some white people and also the scarcity of white clerical and technical workers. Second, there is the principle of the vicious circle. Since there has been but a handful of jobs for trained Negroes in the mainstream of American life, Negroes have not had the incentive to prepare themselves to qualify for these openings.

Third, those who did sacrifice to get the training naturally prepared themselves for positions where there was less resistance, such as teaching in segregated schools or working in Civil Service. Many highly trained Negroes are working in Philadelphia, but they have jobs, often in government, where there is less discrimination and more security. (However, the Federal Government is reducing its personnel and many Negroes will feel these cuts.) Fourth, there is the whole socio-economic situation—low income, poor housing, overcrowding, ill health, family disorganization—which operate powerfully against Negro youth in a highly competitive economy. Even public higher education is not “free” to most Negro youth since they have to work to eat, to buy clothing, and to help support their families. Fifth, Negroes are discriminated against by many schools and colleges. Some of the better secretarial schools in Philadelphia, for example, do not admit Negroes.

Still another factor is the inadequate vocational guidance of Negro youth. One finds, even in 1946, that Negro boys and girls are being dissuaded from preparing for positions requiring advanced education and training. Counselors, white and Negro, are amazed when they are presented with positions known to the Placement Service, or the list of Negroes now teaching in Northern colleges.<sup>3</sup>

---

<sup>2</sup> See Haygood, William C., “Negro Teachers in White Institutions” *Phi Delta Kappan*, XXVIII, (1946), 74-75.

<sup>3</sup> Loescher, Frank S., *op. cit.*, pp. 92-93.

*Responsibilities of the Northern College*

Our colleges have a great chance to contribute to equal opportunity in employment. We can participate in the growing movement to eliminate discrimination. We can inform our students of the minority status of those businesses, industries and other institutions which consider applicants on merit. We can examine the policies and practices of our own college communities in the light of democratic principles and sociological and psychological techniques.

Let me elaborate on each of these three areas.

1. Individually we can further fair employment practices by joining organizations which are interpreting to the community the importance of employment on merit alone. We can support local, state and national Fair Employment Practices legislation.

2. Current information on new trends in employment for minorities should be on the desk of every college personnel officer. Perhaps the Council of Guidance and Personnel Associations could take on this important service. If such information were available one would not find, as I have, placement officers automatically referring Negro graduates to Southern schools and colleges. Instead, if the Negro man or woman had the proper credentials for teaching, he would be referred to Antioch College, Wayne University, William Penn College, the University of Michigan or one of the many other colleges which employ Negro teachers. Similarly persons trained for other occupations could be informed of companies living up to the spirit as well as to the letter of employment on merit. Furthermore, Caucasian teachers would be referred to Fisk, Hampton, Howard, Lincoln, Talladega and other Negro colleges which welcome qualified non-Negro teachers.

3. The college itself can do a great many things both in its own policies and practices and to some extent in the local community.

a) Refer on merit alone. When recruiters call, submit the credentials of all qualified persons, regardless of race or creed. Some offices of the United States Employment Service used excellent techniques for taking job orders and making referrals. The Office of Price Administration in Philadelphia

had a careful program to assist its Negro employees to secure jobs in private industry.

b) Employ in the placement office members of the difficult-to-place groups. One university enrolling thousands of Negro and Jewish students, with a placement department of 20 people, employed not a single Negro or Jewish person. A trained person can assist in several ways. First, such persons through staff meetings and daily social contacts can help to educate the placement staff. Second, they demonstrate to prospective employers where the college stands. Third, they indicate to the students of minority status that the placement office is not just for white Protestants and Catholics.

c) Furnish guidance for the difficult-to-place. This can be done individually and also through groups. Give accurate information on the placement problems of minorities and some of the new trends. Caution: *do not dissuade any person who has the aptitude and interest from preparing for any occupation.* The aim here is to help the individual to appraise himself and his situation realistically.

d) Use local and out-of-town agencies specializing in minority employment problems, such as the Employment and Vocational Bureau, the National Urban League, the American Friends Service Committee, which often have important information and services.<sup>4</sup>

e) Publicize your concern. In your literature addressed to employers indicate that the college has students of various nationalities, religions and races and point out cases of good work adjustment of the different members of the student body.

f) Compare experiences and techniques with other college placement offices. Develop joint approaches to employers.

g) Enlist the interest of sympathetic employers by personal contacts and through personnel organizations. Bring together a group of employers and have an industrialist who practices fair employment share his experiences and know-how.

Throughout the college there is need for a new emphasis on the integration of minority groups. Students of minority

---

<sup>4</sup>The American Friends Service Committee in the near future hopes to make available a listing of many important institutions in which Negroes are employed in non-traditional positions

status on many of our campuses very often do not have a sense of community, a feeling of belonging. Faculty and administration, student organizations and joint faculty-student committees might well study the position of minority groups on the campus and in the town. Negro students, as the *National Survey of the Higher Education of Negroes* demonstrates, in too many of our Northern colleges suffer from a feeling of being outsiders.<sup>3</sup> It is no wonder that every year hundreds of superior Negro young people, and especially girls, go to the segregated southern institutions. Other students—Japanese, Chinese, Indian, Mexican, Jewish, Catholic—often experience a similar type of isolation. Many times these students at Northern colleges do not make use of the personnel services available because of ignorance or a feeling that the special services are only for the majority group. What is worse, because of the stratifying and segregating tendencies on many campuses, students of minority status do not develop their aptitudes and personality traits in a way which would enhance their employability—not to mention their own sense of well-being.

Realizing something of the conditions you labor under in the present emergency, I cannot press you too hard. Nevertheless, this is a period of establishing new patterns. College faculties are rethinking the place and function of higher education in a democratic society. In the years ahead a steadily increasing proportion of college age youth will be attending our institutions.

As I see it, one of the important issues facing the American College Personnel Association is the role of the college community in the occupational integration of minorities. I am certain that here is a place to make an important contribution to a more democratic America.

---

<sup>3</sup> Caliver, Ambrose. *National Survey of the Higher Education of Negroes*. Washington: U. S. Government Printing Office, 1943. Vol. IV.

# GRADATIONS OF DIRECTIVENESS IN COUNSELING

EDWARD S. JONES

Dean of Students, University of Buffalo

THIS paper is an attempt to explore briefly the areas of counseling in a college personnel office with special reference to directiveness (and non-directiveness) in the advising of students. We have all been impressed by the stimulating and enormously vital emphasis of Carl Rogers, Nathaniel Cantor and others who have called attention to the non-directive method.

Briefly I would say that it has led us to the general acceptance of the following conclusions: (1) Most of us have talked too much in counseling young people. We have not stimulated thinking on their part. (2) When we have given advice it has usually not been taken, or taken half-heartedly, because it was never absorbed as a part of the thinking of the student himself. (3) We usually have not taken the view of Max Wertheimer on "productive thinking" in his recent book that education has been over-directive all along the line; that it takes time and special teacher-techniques to encourage student initiative (not mere participation) in solving original problems. Guidance problems are unique and individual par excellence. Better than the classroom they furnish an opportunity for a student's responsible thinking.

However, when we come to apply this point of view in actual practical counseling a few difficulties or questions appear. First of all (and Rogers would, I think, agree) the great majority of conferences in personnel work are directive and informational. We point out schedules, study methods which have been found successful, types of preprofessional requirements, etc. In a recent survey at the University of Buffalo we have observed that about 5 per cent of the actual contacts made with students

involved deep-lying anxieties and conflicts where non-directive techniques are particularly indicated. Of course, in terms of the total time taken in conferences, this runs to between 10 and 15 per cent.

Secondly, granted the value of more of this type of non-directive counseling—for example, in connection with poor study habits—it is very time-consuming. For obvious reasons of budget, relatively few people can be dealt with. The method of brief casual, non-directive “counseling on the hoof,” which Rogers and Wallen refer to, strikes this speaker as fitting and very effective in a few cases, but this casual counseling often is mingled with directiveness.

Thirdly, why is there so much emphasis on “all or none”—either directive or non-directive counseling? Why not various stages along a continuum? To read some of the material it would seem that a spell is cast, that one must not get away from that particular non-directive spell or everything will be lost.

We have two objections to a purely non-directive approach—to the bitter end. First, in a certain proportion of cases counseling just stops; it evaporates into thin air because, as clients say, “I quit going because I got no value out of it.” In an entire industry where several non-directive counselors were installed, the system broke down because it simply was not used enough.

Our other objection is that to confine all of a counselor’s verbalism to “Is that so?” or “You believe that?” or some such non-descriptive responsiveness is artificially non-cooperative. Conversation is not like that, nor are friendly discussions of any sort. Even if there is a general emphasis on student-thinking, or client-centered unravelling, there are so often opportunities to give specific information, to mention a clue, or to illustrate what someone else has done or considered. For this reason, I would suggest several stages of compromise. One may stick to directiveness for a few minutes, then shift over; perhaps, again, one may go in the other direction—from a purely non-directive approach to one that is loaded with specific suggestions or information.

There is merely time for a few illustrations of these stages.

First, there are a few instances of personnel office contacts which are largely routine, and directive in the sense that specific questions are asked—but which often lead beyond into conflict situations.

To begin with, then, there is the orientation conference, which most of us, I take it, carry on as a routine procedure for most students in one form or another. A student is asked to come in to discuss his life on the campus and how everything is getting along. About how many hours a day does he study? Where? What outside work is he doing? From conversation it appears, for example, that all of the studying is done in the library, but that he does not have enough time for it. So far the emphasis is strictly directive and steered. But then comes a break. "It's funny you can't get any studying done in your own home. I wonder why." Suppose the home is quiet, but the man honestly believes that he is a peculiar type of individual who needs a library atmosphere, and not a quiet room by himself. But then he is asked, "Why are you so different?" "Have you really tried other methods long enough?" Certainly this would not be accepted as a pure non-directive start, but from here on it may lead into a good deal of self-analysis, covering two or three return sessions when a student discusses other experiences, largely on a non-directive level.

In the second place, there is a specific problem on which a student wants suggestions, help. Perhaps there are several alternatives that he, the student, may never have thought about. Merely the presenting of alternatives will open up new channels of thinking. The counselor *helps the student to explore*; he does not do all the thinking. For example, a student—a returned veteran—remarks that he cannot concentrate. He tries to study at least a couple of hours but his mind is continually wandering. The personnel counselor helps him to explore. First, there are physical factors—his coffee drinking, his smoking, his lack of exercise. Again there are aptitude tests—his vocabulary and normal reading ability. Then there may be love affairs, or disputes with his parents. Often there are definite clues from his routine responses on a personnel questionnaire blank. These can be probed. To hypothesize

that a man must learn everything himself the hard way—through no suggestions or advice—must discover every alternative possible cause—is like assuming that one must be a criminal to understand crime.

One of the most effective methods of modified non-directiveness is to give *illustrations* of what other students have tried out and found useful. A man, for example, remarks that his grades in final examinations are always at least one notch lower than in recitations and small papers. Why should he always slip down in the big final examination of a course? He is not looking for a rule of thumb, a brief bit of information, but honest clues. He is asked to inspect some of his final examinations carefully; in what types of questions is he weakest? How does he handle long essay questions? Has he ever tried a brief preliminary outlining? Then when the student reports that he always believes in a brief pertinent statement "Because professors don't like to read long statements," he is told about a student who doubled the length of every examination statement he would normally have written—to raise his average grade a whole letter.

In the illustration device the counselor is not taking all the responsibility; it is something which may or may not work. It is worth exploring, and the student is asked to return after the next final examinations to discuss the situation.

Third, there is often a person who needs to have a spade called what it is. An advanced superior student was discovered to be copying material on questionnaires for the Psychological Corporation. He needed a strong jolt: "This is deliberate lying." "It is a serious character blemish." "Where did this get started?" "What other behavior is similar?" These were some questions raised in a series of interviews with the student.

The method of Voltaire's *Zadig*, or of "post-diction" is advocated in such cases. By this I mean one may lead from the present into the past, filling out many details, and in so doing help one to predict, or to forestall the future. A soul-searching analysis of past conduct and motives may be handled sympathetically and quite non-directively.

Fourth, I would like to mention what a recent writer, Hada-



mard, calls "precising" in his book called *The Psychology of Invention in Mathematics*. This, he believes, is the most conscious and rational part of all invention and productive reasoning. By it he means that we talk about a problem enough to really understand it. Oftentimes this is all that is necessary—merely ventilating the whole business. For example, a Public Law 16 (or disability) case who was obviously a neurotic individual, of good ability according to test scores, but a man who had done poorly in two other curricula (and with a very mediocre record in high school) wanted authorization to take up pre-medical training. It seemed desirable to suggest that the student talk about the number of times he had previously made up life decisions, and why he had changed them; also to indicate what patterns of vocational activity he had stuck to for any length of time. It came out that his father needed him badly in a wholesale business just before final examinations. Most of this ventilation of past purposes, feelings, attitudes, was, I believe, largely non-directive; but from there on a fair amount of directiveness entered the picture. The University could not guarantee, or feel reasonably certain, that he would be admitted to Medical School in two or three years. Some compromise program was necessary.

In all of these instances I have merely tried to indicate ways of getting at what Lewin calls the "motoric regions" of the individual, what it is that runs the various engines of the individual, and what quenches the fires which may flare up for a time. Is this not a cooperative professional job, even for therapeutic purposes, to say nothing of analysis? In other words, in line with Dante's allegory, in the purgatorium a man must work out his own purifications, but there should be tests and help available for honest searchers. If not an "angel on his shoulder" a man should be able to benefit by a few clues from friendly advisors.

## FACULTY COUNSELING FOR FRESHMEN

LONZO JONES

Coordinator, Student Personnel Services, Indiana State Teachers College

BEGINNING with the Fall quarter of 1946-47, a system of freshman counseling was inaugurated at the Indiana State Teachers College at Terre Haute. In July, at a meeting with the heads of departments, the plan was explained and each was asked to nominate members of his staff to serve as counselors to an estimated average of approximately twenty freshmen each. In all, forty-eight counselors were nominated and assigned by the Dean of Instruction to counseling duties, in addition to their regular teaching load.

This plan was the second choice among those submitted. A plan to use sixteen counselors, each teaching half-time and counseling half-time, was the preferred plan recommended; but because of a shortage of faculty members and an unprecedented enrollment all were needed for full teaching schedules.

A nine-page mimeographed bulletin for lower division counselors was prepared and a dinner meeting with all of the counselors was called for Sunday evening at the beginning of Freshman Week.

The primary principle of student personnel service upon which the counseling program rested was stated as follows: *"The first task of the college and hence of the Personnel Department is to promote each student's progress toward his own educational goal, at the level of his ability."*

Each Freshman had specified his tentative educational goal on his trial program card filed some weeks in advance of the date of registration. Each was assigned to a faculty counselor who was teaching in the department in which the student's major would fall, and, when possible, with whom the student was enrolled for a class. This was done on the assumption that

such a teacher-counselor would know more about the required courses, the sequence of courses, and the related elective courses which a student majoring in his department would need, than would a counselor assigned at random or from some other department. Likewise, such a counselor knowing the personality and social qualities characteristic of successful persons in the freshman's preferred professional field, could give better guidance when personal adjustments needed to be made. Thus a freshman intending to become a home economics teacher was assigned to a counselor in the Home Economics Department; a prospective coach to a physical education counselor; a prospective elementary teacher to a counselor in the Department of Education; and a pre-medic to a teacher in the science department interested in that field.

Both counselors and freshmen were told that the first standard of achievement by which we would judge a student's "educational progress" was not the requirements for graduation, but his own *level of ability* as revealed on two standardized examinations: (1) *The American Council Psychological Examination* and (2) *The Iowa High School Content Examination*. There are other factors not readily measurable at the time a student enrolls as a freshman which contribute to or detract from his later achievement. Three of these are: (1) his perseverance, industry and application to study as revealed in the number of hours per week which he devotes to study; (2) his motivation as revealed by his intensity of learning or concentration and his willingness to apply improved study methods; and (3) personal factors such as commuting time, outside employment, financial and family worries, and ill-health. However, the two factors, intelligence and previous levels of learning, measured by the examinations referred to above give reliable measures upon which to predict the scholarship index or level of learning for each student.

Errors in examination results (from whatever source) and errors in teachers' marks as well as the factors named above would tend to lessen the accuracy of these predicted indexes. On the other hand, when a student's achieved index varies markedly from these predictions, the variation itself signals

the need for an analysis of the other factors. These may be remedial when brought to his and the counselor's attention.

The counselor was asked to have a get-acquainted interview with each of his counselees during Freshman Days; to arrange a schedule of weekly conferences during the first six weeks; and to have a progress interview when the mid-term grades were available. The freshmen were asked by letter and in the Freshman Week Bulletin and in the orientation lectures to confer with their advisors on the schedule indicated above. A supply of weekly time diaries were furnished the counselors and each freshman was asked to keep one each week during the first six weeks of school and to file it each week with his counselor. This was for the purpose of furnishing an objective record of time used. Keeping such a record helps the student to adjust to the time requirements of college. It also supplies objective data for use later on if a remedial program needs to be worked out.

During the third week of school each counselor had a half-hour conference with the Director of Student Personnel at which time the principles of counseling were reviewed. A work sheet supplying some personal data, percentile ranks and a predicted index for each student was given to the counselor. The student's written biography had been given to the counselors at the beginning of the term.

Again, when the mid-term grades were in and computed each counselor saw the Director of Personnel for another half-hour conference in which the mid-term index for each freshman was compared with the predicted index. Those who varied as much as 10 points below were singled out for diagnostic and remedial treatment. Reduced schedule, reduced employment or outside activities, improved study methods, medical care, counseling on personal problems and attempts at motivation were the main remedial measures suggested.

Finally, at the beginning of the second term, after final grades were recorded, the counselors and the Director again reviewed the student's progress by comparing his earned index with his predicted index. The correlation between the pre-

dicted index and the mid-term index was .61; between the predicted index and, the first term index, .68. The average prediction was 51.4 while the average earned indexes at mid-term and at end-of-term were respectively 56.4 and 57.2.

By comparing the mid-term index with the predicted index for any student it was easy to single out those students who were performing below their own ability levels and to direct remedial measures accordingly. Twenty out of the one hundred students in the sample presented ranked from 6 to 38 points below their predicted index at mid-term with an average of 12.5 points below. Fourteen of the twenty made gains by the end of the term of an average of 9 points each; two remained the same; and four made further losses of an average of 6 points.

I am far from satisfied with the progress we have made, but I believe we are on the right track. Counseling cannot be compulsory and be accepted. Many students solve their own problems and make their own adjustments without requiring or feeling the need for highly specialized counseling. The teaching faculty supply a natural, first-hand contact and with a little direction from the central office can do the screening counseling very satisfactorily.

In our bulletin to the lower division counselors we gave instruction which may well be repeated here:

The skillful counselor will build a mutual acquaintance and understanding through his contacts with the student and the information which the student himself reveals to him. Counseling is not cold-blooded and objective. It is warm-hearted and personal. The counselor is to be the student's best friend and advocate. As such he will *accept* the student as he is. He will come to know the student's plans, objectives and purposes. These, too, he will accept without question in the beginning, later he may wish to refer him to the vocational counseling clinic. He will come to know the student's strong points and weak points. Here both objective data and subjective estimates will enter in. He will give the student facts and information regarding college regulations and will call to his attention the opportunities available in curriculum, extra-curriculum and special service areas. If the student is having trouble he will help to discover the sources and will aid him in working out a plan to remedy it. He will not deceive the student nor dis-

courage him, but will help him to face realistically the demands of respectable college standards to use legitimate adjustments to promote his progress, to accept his own qualifications and limitations, and to seek a desirable substitute objective should he fail to meet college requirements.

The counselor should recognize that he is helping the student to grow toward his own possibilities and not imposing upon him a predetermined outcome. However accurately the counselor may diagnose a difficulty or prescribe a remedy there is still no growth on the student's part until he has gained insight, accepted the situation, and has made his own decisions. This type of counseling does not call for less information, insight and judgment upon the part of the counselor but does require kindness, interest, delayed judgment, and infinite patience as the student acquires his own adjustments and growth.

After all, we are expecting these young people to manage their own affairs. After they are well oriented our obligation should shift to their shoulders. We want counseling but not coddling; we want orientation but not domination of the student; we want self-direction and self-discipline not dependency.

It is our plan to have the same counselors serve the students until the end of the Sophomore year. Then when the student enters the upper division to specialize in his chosen field, place him under the counseling of the Head of the Department in which he is to major, who will induct him into the fellowship of his intended profession.

# DISCIPLINE: A STUDENT COUNSELING APPROACH

JOHN D. FOLEY

Assistant to the Dean of Students, University of Minnesota

## *Introduction*

THAT adolescent misbehavior which, in higher education, we call discipline has been a persisting and perplexing problem in colleges and universities for hundreds of years, yet little appears to have been done in attacking this problem systematically. I do not refer here to scholastic problems of a disciplinary nature, such as cheating and plagiarism, but rather to that out-of-class misconduct which occurs and recurs in each generation of college students. Although a great deal has been written about this phase of student life, few writers have suggested remedial or preventive techniques which are of assistance to the administrator who faces such problems. Since it seems likely that discipline will always be with us because each new generation of college students must repeat the life cycle of the species, it is important to plan systematically to cope with this problem.

There are three points which are important in our orientation to discipline. First, one must differentiate between delinquent behavior and the delinquent individual. Delinquent behavior is often found in an individual who is not characteristically a true delinquent. In the second place, problems which we call disciplinary are problems of adjustment. Social and ethical deviations are symptoms of maladjustment as are problems of scholarship, personal conflict, or vocational choice. Finally, delinquent behavior is a function of the number of laws and regulations which are set up to effect the social control of students.

In an attempt to meet this problem of disciplinary remediation and prevention, a separate bureau was established five and a half years ago in the office of the Dean of Students of

the University of Minnesota. This paper is, in fact, a summary of our experiences. In our program, two psychologists devote their time exclusively to the clinical counseling and rehabilitation of students involved in various kinds of misconduct. The bureau has a suite of offices, a secretary and a separate budget. It is coordinate with the Student Counseling Bureau, the Student Activities Bureau, the Bureau of Veterans' Affairs, the Housing Bureau, the Speech Clinic, and the Bureau of Loans and Scholarships. The two psychologists have had training in the genesis and treatment of behavior problems in addition to training and experience in general student counseling. It is believed that the combination of such training best prepares them for the specialized work of rehabilitating students who commit delinquent acts.

The psychological and social context in which student offenders are counseled is extremely important. Disciplinary problems tend to be associated with morale problems. All of the taboos and social sanctions which are related to misconduct and which give it unacceptable overtones and undertones create the need for a specialized approach to this type of social and personal adjustment rehabilitation. The disposition and counseling of each case are prescribed by the disciplinary counselors. All cases are subsequently reviewed by the All-University Disciplinary Committee. Although the counselor may refer a student to other personnel workers or to psychiatrists for additional specialized treatment, he retains his central role as the "behavior teacher." Services which are provided through referrals, such as remedial reading or financial aid, are coordinated and integrated into the individual rehabilitation counseling program. They are important in themselves but nonetheless they are adjuncts to the central feature of the program—the rehabilitation counseling. This situation differs from the one in which a student is disciplined administratively and is then referred to a counselor for psychological services. This point will be discussed in greater detail later.

#### *Disciplinary Problems of College Students*

In 1945-46, the University of Minnesota received complaints about the behavior of 163 men and 146 women. Twelve



men and 19 women students had more than one complaint made about their behavior. This recidivism during the year was 7% and 13% respectively. Counseling services or administrative action occurred in 66 other instances involving men, and in 13 others involving women, all of whom had been the subject of complaints in years previous to 1945-46. The total case load was 419 in a student population of about 20,000.<sup>1</sup> Complaints about student organizations are not included in this figure. A total of 3620 contacts were initiated with these students, their counselors, faculty members, parents and others in an effort to understand the problem behavior of these students and to effect their rehabilitation. This figure does not include telephone calls to students requesting appointments but only those contacts which were part of the case work.

Slightly less than 41% of the complaints made about male students were lodged against veterans. Only 2 complaints were made about women student veterans. Space does not permit a detailed discussion of the disciplinary problems of veterans but certain generalizations about them may be made. These students are older than the typical college student and have better integrated habit patterns. Not only are they more mature and, at least superficially, more sophisticated, but they appear to have broader horizons, both with respect to education and with respect to social and recreational habits. In counseling student veterans who have been delinquent, the circumstances of their war experiences are taken into consideration, but this adaptation to them is no greater than it would be for any other student who is counseled according to his particular developmental history.

Student misconduct at Minnesota is classified in six categories. Additional statistics for the year 1945-46 are given together with descriptions of the classification:

1. *Financial irregularity*—non-payment of bills or rent; writing worthless checks—13.1%.

---

<sup>1</sup> Case load is here defined as the sum total of students-in-complaint situations. A student who gets into trouble twice is counted twice in the case load because two separate problems are presented to the counselor. If two or more students are involved in one disciplinary situation, the case load total will be increased by the number of students in the situation since each presents a separate counseling problem.

2. *Sex misconduct*--all kinds of sexual irregularities including pathological types--7.4 .
3. *Disorderly conduct*--more serious types of misbehavior, such as gambling, drunkenness and brawling--13.6%.
4. *Theft*--stealing, embezzlement, and forgery--12%.
5. *Misconduct*--disputes involving interpersonal relationships; infractions of housing, and other university regulations. These are considered less serious than those classified as disorderly conduct--37.5%.
6. *Miscellaneous*--misuse of student privileges or university facilities (library, athletic, student union); dishonesty; lying; alteration of official records; and unusual kinds of situations such as pouring acid in ballot boxes during a campus election. These are usually minor offenses. None of these contributes a substantial proportion of complaints. They accumulate, however, to 16.4% of all complaints.

Most misconduct results from one of three sets of conditions. These may be called inadequate development, neurotic development, and lack of information or understanding. Inadequate social or emotional development refers to gaps in learning how to behave properly. This may be a function of psychological dependence, a careless and uncritical tendency to do what others do regardless of its acceptability, or the failure to "think through" outcomes of behavior. A related phenomenon is pathological development--the "normal" social and emotional development of an individual who grows up in an abnormal community, one which is in itself deviate in terms of our typical customs and mores. Rehabilitation under such circumstances is extremely difficult to achieve since it involves a complete re-structuring of social and ethical attitudes.

A second condition which may give rise to misconduct is neurotic development. Misconduct may be an expression of neurotic personality. An example of such misbehavior is that kind of stealing which serves to reduce neurotic tensions as differentiated from kleptomania. These individuals are probably responsible for many thefts which are never solved because no one recognizes their techniques for reducing tension. Healy has discussed them in great detail.<sup>2</sup> Most psychologists and psychiatrists who have dealt clinically with behavior problem cases have had experience with misconduct originating in this

<sup>2</sup> Healy, Wm. *Mental Conflicts and Misconduct*. Boston: Little, Brown, and Company, 1917. pp. xi + 330.

fashion. Our clinical counselors are trained to utilize appropriate clinical methods in the treatment of such cases.

Many cases of misconduct arise because of ignorance or misunderstanding on the part of the student. It is not uncommon for the adolescent entering the complex environment of the college to stub his toe on some regulation about which he has never been informed or which he does not understand adequately. Even though a lecture or a rebuke might bring about appropriate understanding by these students, their disciplinary situations may more profitably be turned into teaching sessions where attitudes can be explored and where other areas of ignorance can be corrected.

Sometimes misconduct results when a student is experiencing a psychotic episode. Occasionally a delinquent is a psychopath. Treatment of such individuals is not undertaken by the disciplinary counselors in such instances but psychiatric services may restore such students to better levels of adjustment.

#### *The Minnesota Disciplinary Program*

*Philosophy.*—As in many other schools, the student personnel point of view has been adopted as a basic philosophical approach to discipline. We have instituted practices which are consistent with this philosophy and which provide the student with maximum opportunities for learning more acceptable adjustment techniques than those which have brought him into conflict. Students are not mollicoddled but are brought face to face with their problems realistically so that they may learn more effective ways of everyday social and ethical living. In this context, the disciplinary situation becomes a learning opportunity, as does any properly structured counseling situation. We apply the principles of the psychology of learning to teach the student that which he should have learned before coming to college. Many teachers who have not dealt with discipline often do not realize that a large number of students have never experienced an ethical awakening. It is consistent with the student personnel philosophy that the college undertake this task whenever necessary, in order to prepare students for the world in which they must live.

The adoption of the student personnel philosophy does not imply that all students must be kept in college. It is not always possible or feasible to rehabilitate a student on the campus, but such cases are rare. The chronic dipsomaniac and the aggressive sexual psychopath with overt manifestations of sadism ordinarily cannot be rehabilitated in college. Nevertheless, a student cannot be re-educated by depriving him of the counseling and educational facilities of the university. Hawkes and Hawkes have stated this clearly and forcefully:

The student has been accepted by the college, he has been called by a new name, because it was assumed that he was worth educating. To throw him out at the first sign of immaturity or of failure, is nothing less than shirking duty. It is easy to conceal such incompetence by calling it maintenance of standards, but such an alibi deceives no one except those who use it.<sup>3</sup>

Credit penalties, fines, lectures, and scoldings are not used in our program. The situation is not publicized to make an example of the student. A student may be punished, his privileges may be restricted, or he may be dismissed from college, but the student understands that this is appropriate to his rehabilitation. When a student is punished, his punishment does not fit the crime; it is tailored to fit the individual-in-the-situation. If an educational point of view is applied to the disciplinary program, the student usually enters willingly into the prescribed rehabilitation program and accepts the consequences of his misbehavior. The student who is dropped may still utilize counseling services to carry out his relearning process.

There are certain larger aspects of college administration which may modify circumstances in a particular disciplinary case. Public relations must be maintained. The morale of other students who know about the transgression must not suffer. In many institutions, too much emphasis is placed on misconduct as such. Unless delinquent behavior is to be considered as a symptom of maladjustment, the public act itself may be judged without reference to the student's potentialities. Furthermore, discipline tends to become an isolated and secretive, if not furtive, area of student life.

<sup>3</sup> Hawkes, Herbert E. and Hawkes, Anna L. Rose. *Through a Dean's Open Door*. New York: McGraw-Hill Book Company, 1944. pp. 194-5.

Our experience at Minnesota indicates that these difficulties can be surmounted by giving professional status to the program and integrating it with other aspects of the personnel program. Counselors in other bureaus accept it as a professional service. The administration and the student body also accept it.

*Structure.*—The disciplinary counselors receive and investigate complaints, interview the students and formulate and carry out the therapeutic program. The Dean of Students may be consulted in a particular case, or it may be referred to the All-University Disciplinary Committee. This Committee is composed of faculty members from Law, Medicine, Sociology, Psychology, and Education in the belief that these sciences have the most to contribute in effecting adequate rehabilitation of delinquents. The Committee, patient and sympathetic, is effective in securing the cooperation of recalcitrant individuals and in preparing them for rehabilitation. In many instances, committee meetings are of a therapeutic nature. The attitudes of students change perceptibly as the offenders realize that the group is trying to understand and help them rather than trying to hang something on them.

When it seems unlikely that the student can be rehabilitated within the framework of available campus facilities, the Committee may recommend suspension or dismissal to the President. Only the President can dismiss a student permanently or suspend him for more than one year.

Only a small proportion of students are presented directly to the Committee; the greater proportion of these are retained in school. The Committee may formulate recommendations for the counselors to carry out or it may give the counselors *carte blanche* to act as they see fit.

All cases are reported periodically to the Committee for review. This provides a double check on the judgment of the counselors. Students are advised that they may appeal to the Committee if they are dissatisfied with the treatment which the counselor has prescribed. This gives the student confidence that he is being treated fairly and usually insures his cooperation. Few avail themselves of this opportunity.

Students may appeal from committee decisions to the President who appoints a Special Review Committee. Through this

series of appeals and reviews, the interests of the student are protected against capricious or arbitrary judgments.

### *Procedures in Disciplinary Counseling*

*Investigation of Complaints.*—All complaints are referred to the disciplinary counselor who interviews the complainant about the details of the misconduct. The social and psychological dynamics of the situation are ascertained and an evaluation is made of the sincerity and integrity of the complaining witness. This prevents the student from being victimized through vague and unreliable complaints. At the same time, some understanding is gained of the magnitude of the problem. If the complaint is not a valid one, the situation is studied to determine why the conflict arose. If, for example, the complaint arose in a rooming-house, the situation is evaluated to determine whether the student is socially awkward and in need of counseling, or whether the householder misunderstands her role in student life and needs further indoctrination.

*Collection of Information.*—All pertinent information about the offender is then collected. This includes his scholastic record, his entrance test scores and other admissions data, and reports from counselors, advisors, and other university or non-university personnel who have had contacts with the student. When residence reports can be secured, they are frequently extremely valuable because in his residence the student is most likely to behave in his natural way.

This information is obtained and studied before the student is interviewed, to gain further preliminary insight into his personal and social adjustment. The student is told that a complaint has been made about his behavior and that the counselor's job is to help him to resolve the situation. The counseling process begins at this point. Usually the student will give his account of the story in a frank and above-board manner. His account may differ or may agree with that of the complainant. If he begins to lie or to stall, the interview is terminated and he is asked to think the matter over for a day or two and then to return for another appointment. This usually secures the cooperation of students who may be frightened or

who may try to bluff their way through the interview. Since students are often upset by the difficulty, they are usually anxious to see the problem cleared up.

The complainant is not identified to the student even though he may know who has reported him. The purpose of this is two-fold. First, the student is encouraged to look forward to the counseling process and to evaluate his behavior objectively in order to secure better future adjustment. Since the past cannot be undone, it is better to structure the situation so that he may learn something constructive from it, not merely how he failed to cover up his misdeeds. In the second place, a student occasionally wants to revenge himself upon the complainant. Even though the latter may have had mixed motives in reporting the incident, he should not be subjected to indignities because of this. Complaints are used to identify students who are not getting along harmoniously and, therefore, have more value than the immediate resolution of the conflict. The interview, therefore, is structured to show the student the university's interest in him. Reconciliation between the disputants usually follows.

*Formulation of Rehabilitation Program.*—Once the facts have been established and the two stories synthesized, the clinician formulates a program designed to re-educate the student insofar as that is necessary. This may consist of anything from a simple counseling and explanatory interview to a prolonged program of intensive attitude or behavioral therapy. Services of any personnel bureau in the Office of the Dean of Students may be utilized in the treatment program. Control of the case, however, is centralized in the hands of the disciplinary counselor.

Some persons believe that such specialization in disciplinary counseling is unnecessary. They ask why an administrator may not take disciplinary action against a student and then refer him to a counselor or a psychiatrist for therapy.

Several cogent reasons for this may be mentioned briefly. In the first place, forced referrals are ineffective if the student is not in a state of readiness for change. Furthermore, the dual relationship of the student to two individuals creates bifurcation rather than integration in the rehabilitation program.

In the second place, the student may avoid coming to grips with his central problem of adjustment. He may repress the first part of the situation, the punitive part, and work out a superficial adjustment with a counselor.

The third difficulty which may be encountered is that of communication. Investigatory interviews elicit certain subtle personality factors and attitudes which influence the counseling process. These aid in determining the appropriate techniques for teaching the student. Our language for describing attitudes and motivations is somewhat less than perfect and it is difficult enough for a clinician to orient another counselor toward all of the pertinent personality factors in a given case. An administrator who is not trained as a psychologist may miss such factors altogether. The counselor to whom the student is referred may secure the background of the problem yet find it very difficult to fill in the foreground.

Perhaps the most important argument against such an alternative concerns the nature of a therapeutic process. Re-education should begin in the first interview and proceed throughout all contacts with the student. A considerable degree of change should have taken place by the time the resolution of the conflict is reached. Disposition, settlement, or action—whatever takes place to satisfy the complainant or to put an end to persistent delinquency—is merely a phase in the total rehabilitation progress, not a major landmark. To the complainant, this may "close the case." But if the matter is not a simple one, this is merely a transitional stage in the learning process. Furthermore, the factor of motivation is important to the therapeutic process. Psychologically speaking, the student's defenses have been broken down in the earlier and perhaps conflictive part of the process. Now that his weaknesses have been aired, he has little to hide from the disciplinary counselor. The fact that he has been caught is the most potent weapon for motivating him to change his behavior to more appropriate standards or levels. If he continues the educational process with this same individual, he can be more relaxed and cooperative. He will not be traumatized by re-experiencing the conflict in a new relationship with another individual.



Healy has emphasized the necessity for skillful, expert handling from the very beginning of a case:

One of the leading points in method is concerned with the attitude of the one who would, for purpose of therapy, assail these problems with the individual offender. . . . The attitude of the analyst should be the antithesis of one who settles the case without full inquiry, or who in an off-hand way gives an opinion of why a certain misconduct has occurred. The approach should be such that the delinquent himself feels that the inquiry into foundations is born of the desire to help.<sup>4</sup>

It is argued that counselors in this field should have special training above and beyond generic counseling training. This work is specialized as is that of the non-directive therapist, the group therapist, the remedial reading specialist, or the speech clinician. Although all of these personnel workers have a common core of psychological preparation, they begin, at advanced levels of study, to specialize as do medical men and social workers. The typical counselor cannot, therefore, be considered qualified to undertake such therapy even though he may understand the general principles involved. Both specialized training and clinical experience are necessary for deft handling of discipline cases.

Finally, punishment prior to insight is like the cart before the horse. When the student has achieved insight into his behavioral dynamics, punishment is unnecessary since he will be better able to behave in acceptable ways. Experimental evidence indicates that punishment is relatively ineffective both as a deterrent and as a corrective. Consequently its use should be minimized in educational institutions.

*Other Treatment Considerations.*—Treatment can be personalized as punishment can never be. It must accomplish justice for the injured person and it must motivate and stimulate the student to re-think—if necessary, to re-cast—his attitudes and beliefs which are inappropriate to the educational setting. In addition to environmental manipulation, psychotherapy, group therapy, follow-up, and other counseling and personnel techniques,<sup>5</sup> the counselor may place restrictions on

<sup>4</sup> Healy, *op. cit.*, pp. 57-59.

<sup>5</sup> Rogers, Carl R. *The Clinical Treatment of the Problem Child*. Cambridge: Houghton Mifflin Company, 1939.

certain privileges because the student has failed to show that he merits them. It may even be necessary to recommend suspension or permanent dismissal.

Where permanent dismissal is indicated, the student is made to understand that this step is taken because it seems to be the only appropriate course of action. He learns that such a step is taken because he is not yet ready to fit into the university community and that the counseling and teaching facilities at the university are insufficient to prepare him for citizenship at his present level of development. He is told that he may apply for readmission whenever he believes he has learned to assume his responsibilities. In this sense, permanent dismissal from the University of Minnesota is never considered permanent. During the past five years, a dozen students have been expelled. Half a dozen students, previously expelled, have been readmitted.

In this frame of reference the behavioral act itself is considered only as a symptom of possible maladjustment. The various kinds of misconduct therefore assume different values for the counselor than they would under some other program. Minor acts may identify severely maladjusted students and the converse may also be true.<sup>6</sup> In terms of treatment outcomes, for example, sex offenders may have better prognosis than those students who steal because of a neurotic disposition. When all factors are taken into consideration and balanced in the treatment program, the student gains not only insight but also "outsight." Should he leave school as a consequence of his misbehavior, he is invited to utilize the services of the counselor at any time he wishes to do so. The majority of those few who are dismissed take this offer literally.

In 1945-46, the disciplinary counselor had 38 consultations with former students who, at one time or another, had met him through disciplinary situations. The majority of them had not

---

<sup>6</sup> Recently four students were apprehended while gambling for small stakes. Two were found to have marked hysterical symptoms; a third had serious personal and family difficulties; the fourth had no adjustment problems. Instead of punishing the students, the counselor treated the problems discovered during the initial contacts. The use of certain facilities were restricted but emphasis was placed upon therapy. After the first interview, no further mention was made about the complaint. It had served its purpose by identifying three students needing mental hygiene counseling.

been dismissed but had graduated or entered the armed forces and were seeking counsel about their future plans.

*Disposition of Cases.*—In 1945-46, 3 students were permanently dismissed, 5 were suspended for varying periods of time, and permanent dismissals of 2 former students were rescinded. One student withdrew to avoid the consequences of his sex misconduct and 4 former students were denied readmission on the grounds that they were not yet in a state of readiness to fit into the university community.

Fifty-eight students were placed on disciplinary probation and received extensive counseling services from the disciplinary counselor, the psychiatrist, or both. All but 4 of these students were removed from disciplinary probation during the year. One hundred and twenty-four students were counseled intensively over a relatively long period of time. One hundred and seventy-four students were counseled in one or two interviews. In some of these cases privileges were restricted but all students were counseled. Emphasis was placed on the latter rather than on the former.

Thirty students were found to be not guilty of the charges which had been made. Many of these exhibited some form of social or ethical awkwardness and were counseled with respect to their appropriate weaknesses. In 27 cases, no final disposition was made. There are two reasons for this unsatisfactory termination of complaints. In most of the cases the complaints concerned theft or property damage by students who could not be identified. Investigations were made but the perpetrators were not discovered. In a few cases the students left school before the complaints were made. No indications of maladjustment were uncovered and the complaints were of such minor consequences that no further effort was made to reach these students. Most of them had graduated and left the city.

### *Prevention of Delinquency*

The authority of the All-University Disciplinary Committee goes beyond that of mere jurisdiction over discipline cases. The Committee is charged with the responsibility for investigating those circumstances which give rise to misconduct and for recommending policies designed to correct them. This gives

broad scope to the Committee's function and permits an integration of committee work with other phases of the personnel and non-academic program of the university.

The remediation of such conditions in order to prevent the occurrence of subsequent misbehavior calls for a broad attack upon the problem. Since a large proportion of misconduct arises in housing situations or in group activities, the work of the disciplinary counselor is constantly coordinated with that of the Housing Bureau and the Student Activities Bureau. A student residence may be thought of as the analogue of the foster home which is used in juvenile delinquency case work. Certain aspects of the student's foster home may be utilized in his readjustment. As a generalization of this concept, the Housing Bureau places each student in accordance with his needs.

Similarly, group activities provide normal social controls and restraints which tend to keep students from getting into serious difficulty. Activity counselors help groups to maintain balanced programs which meet the social needs of college students. Strong organizations are antithetical to serious misconduct. In-service training programs are conducted for fraternity and dormitory counselors, for sorority housemothers, and for private householders who live in the university area. In regular meetings held throughout the year these personnel are systematically educated concerning university policies, educational philosophy, and student adjustment. These people help us to identify maladjusted students at an early period of their university residence so that remedial steps may be taken while the problems are still minor.

We plan to evaluate our program in the near future. Criteria of success, other than the absence of recidivism, must be developed. It may be necessary to follow our clients into their post-college careers. One real test of the program, however, is the criterion set forth in the student personnel philosophy—the obligation to emphasize the development of the student as a person rather than his intellectual training alone. Has the university utilized its facilities in educating the student? Has the university accepted and met its responsibility to make its students fit for the world? This criterion we have attempted to meet.

## THE APPLICATION OF PERSONNEL METHODS TO UNIVERSITY HOUSING PROCEDURES

B. JAMES BORRESON

Director, Student Housing Bureau, University of Minnesota

DURING the past 18 months the college scene has undergone a radical evolution. New and widely dispersed age groups have replaced the relatively homogeneous student body of pre-war days. Enrollments have reached a flood tide never before anticipated. Most educational institutions face extreme shortages of faculty and classroom space. Long-standing student interest patterns have been displaced for the new. Perhaps the most spectacular of all problems facing college administrators and personnel workers during this transition period has been the shortage of housing. Certainly no other sociological problem area has as lengthy an ideology so misunderstood by so many.

Following the first World War, problems identical with those already noted existed on a lesser scale. In answer to that challenge, a new educational philosophy was raised which was eventually to separate American education from its Germanic fountain-head. The resultant novelty became the commonplaces in higher education today, but as a part of this wide revision the question of housing received only a brief and perfunctory treatment.

A repetition of this process is now occurring and new formulations are already evident, particularly in general education and student personnel work. That such changes are evident already is an indication of the magnitude of the educational crisis fostered by the second World War. The application of broad personnel procedures to student housing is but one aspect of this accelerated development. Probably the immediate pressure of the housing shortage has served as a catalyst for change.

Without such pressure it would have been difficult to abandon some of the excessive emphasis on physical facilities, hours regulations, manners and morals for a more sound approach. The shortage and its consequences cleared away some of the prevalent symptomatic methodology and provided the stimulus for a somewhat different means of meeting the long-standing problem of educational (*ergo*, personnel) use of student environment. At Minnesota the Student Housing Bureau, a personnel arm of the Dean of Students Office, has drawn freely from the applied and academic fields of sociology, psychology, and social work in the formulation of this service program.

An ecological summary is essential to the understanding of the few areas of development covered here. The University of Minnesota is located midway between Minneapolis and St. Paul in a metropolitan area of somewhat less than a million persons. We are bounded on three sides by light or heavy industry, and on the fourth, by the Mississippi River. Between the University proper and these boundaries is a limited residential section which, without the presence of the University, would resemble a zone in degenerative transition. The existence of the University has retarded but not stopped this process. The large homes of a former day have been converted to rooming-houses and those portions of this limited district geographically most remote from the University show the spotty beginnings of a slum area. Strangely enough, the University's former policy of disapproving of very few of these rooming-houses has effectively eliminated the natural replacement process. Excessive land values are maintained by their permanent rental value. A survey of the area in 1943 indicated a maximum rental capacity of little more than 3,000.

Prior to the war the University met student housing demands through its three dormitories and approximately 500 rooming-houses located in the immediate residential area. In 1941 E. G. Williamson undertook the first analytical study on the residential status of some 15,000 students. At that time, roughly 60% of the student body resided with their parents in Minneapolis or St. Paul. Some 40% were classified as non-Twin City residents. Approximately one-half of the non-

metropolitan group were housed in rooming-houses. The remainder were distributed through dormitories, fraternities, and sororities.

General economic recovery accelerated by the advent of war was reflected by an alteration in the ratio long maintained between Twin City and non-Twin City students. By 1945 the relative position of each group was reversed. Only 40% of the student population were then residents of Minneapolis or St. Paul. Although the total enrollment was considerably below the pre-war figure, those in need of housing exceeded the maximum housing capacity of 1941. Our housing shortage began one year prior to the enrollment of 27,000 full-time students.

The transition of the Student Housing Bureau from a listing agency to a coordinated part of the Minnesota personnel team occurred in this frame of reference.

Although the housing of students is historically one of the first service functions undertaken by the college, it has been the last to mature. Even before the war overwhelming emphasis was placed upon providing adequate physical structures to meet the room and board needs of the student population. Any personnel objects above and beyond this service were hazy in formulation and primarily restricted to the verbal level. A methodology for establishing and controlling directed non-classroom learning situations, therapeutic or social, was limited or non-existent. Were one to propose the adoption of our present weighted value system to the social worker engaged in foster home placement the reception would be poor indeed. Recent research on problems of atmosphere, leadership, and integrative behavior further demonstrate the inadequacy of our present objectives if living unit goals are ever to be linked to the goals of general education. Time distribution alone would indicate that much of student learning, particularly social and emotional learning, takes place outside of the classroom. Present non-utilization of the educational opportunity inherent in the living unit is unfortunate but follows naturally from the present emphasis upon physical facilities. If this limitation continues, placement criteria, the early identification of adjustment problems, the utilization of resources other than the

counselor's office for a unified and broadly based therapy, and almost ideal opportunities for socio-metric research will be by-passed.

No one will argue with the premise that an *unsatisfactory* physical environment deters maximum utilization of college opportunities, academic and social, and that University construction in this area should be expanded to meet minimum living needs of students and faculty, married and single, as they exist. But to hold, either (1) that dormitory structure must precede personnel work, or (2) that physical facilities exercise the same environmental effect through all quality and type ranges is, to my way of thinking, fallacious.

Among college administrators there is a general tendency to feel that the main obligation of the college has been fulfilled when adequate dormitories are provided despite the fact that on an individual level those same social and personal adjustment problems that existed before continue. Some of this effect is probably due to several early studies which found a positive relationship between academic grades and types of living arrangements. A confusion of correlation-causation led to a superficial treatment of the important factors of selectivity, socio-economic status, and atmosphere inherent in such projects.

Many personnel workers have tacitly accepted the conclusion that university-owned facilities are the only types of housing units suitable to an extensive personnel program. The validity of these general points is sustained by the literature in this area. Although meager, what exists is concerned almost exclusively with one phase or another of dormitory operation. If the number of students involved in types of housing units is any criterion of comparative need the reverse should be true. Klein's early survey of 69 land grant colleges found only 15% of 136,000 students living in residence halls. Under the pressure of current enrollments this unbalanced ratio between dormitory and non-dormitory students will continue to exist on many campuses for years to come.

At Minnesota, students residing in units other than dormitories constituted 75% of the non-metropolitan student popu-



lation. As a result of their primarily rural background and their dispersion from the campus these students were both ill-informed and reticent to participate in the social program and personnel services made available to the rest of the student body. The relationship between students residing in rooming-houses and those in other residence classifications has been a repeated argument in campus political activities. Virtually every campus political campaign has involved some measure of programming for this forgotten class.

This is the background to the philosophy of providing those personnel and counseling services which should be available under a broadly conceived dormitory program to students in rooming-houses through matched placement, field work, and social programming.

The past two years of shortage has seen an increasing number of students seek the services of the Housing Bureau. The unit interview load in 1944-45 was 4,093; in 1945-46, 16,144. In 1946-47 an estimated 21,000 will be interviewed.

The initial interview and all subsequent interviews are recorded on a numbered housing personnel card which serves as the student case record for one academic year. Each such card covers a wide range of information including all previous placements, personal disabilities, confirmed adjustment problems, financial status, age, college course, housing needs, and other related factors.

If the student has been the subject of a householder complaint, this complaint is classified and posted to his housing personnel card. As data accumulate this card may serve as a gross diagnostic instrument for the interviewer, a longitudinal history on an adjustment area for the clinician, and a reliable summary of placement data. Such information enables interviewers to identify problems of adjustment where the symptom is repeated minor anti-social behavior prior to a major delinquent act. Through referral, counseling may begin at an earlier effective date and environmental conditions may be altered to fit the desires of the clinician. The minimum that can be expected from such a service is student placement which is not detrimental to the counseling process. Therapeutically

such an aid is subject to the limitation of age, the nature of the delinquent act, or precipitating factor, and developmental theory. Two simplified situations will serve as examples of this preventive use of the placement procedure. Identified psychopaths are not placed in graduate houses where restrictions are virtually non-existent, nor are social recluses placed in private homes in the suburbs.

Before considering actual placement procedures, mention should be made of one further interview function. At present the Housing Bureau is one of the first agencies where person-to-person contact between the personnel worker and the student or parent takes place. The size of the University and the present accelerated pace in all of its activities are often equated with the term "impersonal" in the minds of parents and students. We have found that the tension which naturally accompanies the registration process and anticipated housing difficulties may be relieved by re-structuring the university setting in a more personal light. At the beginning of the academic year this orientation process is one of the major phases of the interview.

Placement procedures vary according to the classification given to each applicant at the conclusion of the interview. Four major categories are used.

Students with psychological problems identified either through the interview or through referral from another campus personnel agency are separated from the main group of those interviewed. Speech clinic cases, discipline problems, students with difficult home adjustments and students whose personal or social adjustment may be improved by a change in environment are included in this category. Cripples, diabetics, spastics, and students with physical disabilities are also designated for specialized placement. Action on the application of these individuals is delayed until full information from all campus personnel sources may be collected. As part of the routine placement process for members of this group, householder case histories are reviewed and individuals are placed with those householders who empirically have aided students in overcoming roughly similar problems. The security offered by an

interested family in a private home has proved superior in most instances to the atmosphere of a regular rooming-house. Following such a review the householder receives an abbreviated and simplified statement of the problem and is advised of the methods most useful in dealing with anticipated behavior or adjustment problems. Regular contacts with these householders are maintained through field work.

Students whose cultural backgrounds differ from the American pattern also receive specialized assistance in placement. Although placement criteria differ for these students the process is similar to that used for the physically and psychologically handicapped. To supplement this program three privately owned rooming-houses have been converted on an experimental basis to International Houses. Supervision and placement in these homes rests entirely with the Student Housing Bureau and membership is restricted in the following manner: No two foreign students with the same cultural background may be placed in a single unit, and one-half of the tenants must be American students. Placement in such units follows a more rigorous and lengthy interview than that routinely practiced. A limited testing program for determining initial attitudes and attitude changes is under way in these units.

Most students are classified as regular placements. Their needs are matched with those vacancies which best meet their demands. Each listing rejected and the reason for the rejection is posted to the student's housing personnel card. Final placement is confirmed by a telephone conversation with the householder. If the reasons for the rejection involve the suitability of the householder or the house, the case history folder is checked for confirmation. If the data available are inadequate a field call is made.

Severe neurotic or psychotic adjustment problems are transferred to a psychiatrist or senior counselor. Placement is delayed indefinitely until a recommendation is received.

It is obvious that such a placement program requires more than the usual cooperation with and control over private rooming-house operators. A further requisite is a detailed knowledge of each householder, the history of her relationship with

students, and her weak and strong points from a personnel point of view. The householder is, in fact, the keystone upon which the program is built. Although every effort is made to convince householders of the advantages which will accrue to themselves and to students through a personnel treatment of placement, occasional resort to police powers is necessary. Each student, married or unmarried, graduate or undergraduate, is required to live in an approved residence. If house or householder standards are below the acceptable minimum, approval as a student residence may be withheld or withdrawn. Although this economic weapon is more widely exercised today than ever before, it remains a procedure of last resort.

There are now over 3,000 rooming- and apartment-house operators renting to students compared with less than 500 in the early months of 1945-46. To most of these persons renting to students is a new and unique experience. The long-standing hostilities and traditions of independence common toward any agency that has been identified with police powers is non-existent among these new householders. This has been a boom in educating them to a personnel point of view. They are further interested in the protection that their cooperation with this office offers them as landlords.

In developing the householder aspects of this program four instruments have been effectively used. Field work, in reality social work on a new operational level, is the most important. Four full-time professional social workers are employed and an in-service training program is offered to alter, when necessary, standard social case work techniques so they may be applied to a heterogeneous group ranging from one extreme to the other in educational, cultural, and socio-economic backgrounds. Each such field worker is assigned one of four geographic districts and householders and students within this area are her potential and actual "clients." This roving personnel service is a significant departure from the usual office setting for college personnel activities.

In conducting an educational interview with a new householder, a field worker outlines the part to be played by her in the over-all personnel program. University personnel resources

are briefly summarized, along with methods of securing student cooperation in the observance of general regulations, the behavioral significance of rules violations, methods of identifying social isolates, and a point of view on the range of student living patterns and wherein and why these differ from those of non-student tenants. With the completion of the interview the householder signs a general policy statement outlining the responsibilities of an approved housemother.

If the householder has had previous experience in renting to students a portion of the interview is devoted to specific student problems within her own home. Instances of poor academic, personal or social adjustment are noted for further investigation. Intra-house organization is encouraged and the householder training program and Student Rooming House Council explained.

Immediately following the interview each house and householder is assigned a rating on the basis of a 150-item Rooming House Rating Scale now in the process of validation. This instrument contains three sub-scales with relatively low inter-correlations. If the total score falls within the approved range a detailed inspection of physical facilities is made by the University Health Service. These evaluations and ratings are posted to the Housing Information Cards used by interviewers in matching placement with need. Such cards are the summary histories of materials gathered from all sources on each house for a two-year period. The original information becomes part of the permanent householder case history folder. Should the field worker find the householder qualified and willing to assume a part in one of the experimental programs this judgment is reported, evaluated, and if approved, an additional field call is made to complete such arrangements. A routine householder interview occupies from 40 to 60 minutes.

Field calls are also made in response to householder or student complaints. Complaints may vary in seriousness from incompatible personalities to a full-fledged war of attrition between student tenants and landlords. As is true in most verbal evidence these complaints carry an effective charge and are symptomatic rather than dynamic. Adjudication will vary

from situation to situation depending on the degree of structuring, the amount of emotional involvement, the available or possible solutions, and the prognosis for individual and group adjustment within the same environment. Field workers encourage and attend individual house meetings between the householder and her student tenants. When possible some measure of student self-government is established but staff limitations make group work difficult. The main function of the field worker in such meetings is the development of a permissive atmosphere for a more objective discussion of disagreement and controlling the release mechanisms that are a natural concomitant to the particular setting.

In working with students an anti-symptomatic approach to their complaints has paid dividends. As would be expected, scholarship, social inadequacies, and finance are among the underlying and real precipitating causes of disagreement. The complaint as a readily available point of reference makes referral to other personnel agencies simpler than would normally be expected.

Two types of cumulative records are kept on students. Both are excellent resources for the identification of counseling problems and one serves the secondary purpose of pointing up householder inadequacies. The first of these records is the tabulation and posting of all householder and student complaints; the second, the number of residence transfers made by an individual student during an academic year. The latter, when taken as a cumulative record, may show rather startling results. Certainly, in some instances, more than the normal moving about is a reflection of poor adjustment to the college environment.

Field work is supplemented by a householder training program analogous to that in operation for regular dormitory and fraternity counselors. The personnel role of the approved householder, the personnel resources of the University, property improvement, and home management are among the topics included in the monthly series of lectures and discussions. Once a month a Householder Newsletter is issued by the Housing Bureau. This leaflet summarizes basic policies and previous

training program topics. Local gossip and the designation of a house-of-the-month are included as interest-drawing items.

Although the student organization made up of tenants of rooming-houses and private homes is one of the more interesting aspects of the total program, an adequate discussion of this threshold group to student activities would consume far more time than is available. I hope it will suffice to say that the Student Rooming-House Council serves as a counter to the social isolation usually taken as synonymous with residence in a rooming-house or private home.

I have attempted to describe briefly a few of the personnel techniques and procedures applied to student housing at Minnesota. The program relies on an integrated series of services and a new application of the already established principles of social case work, social group work and counseling. I believe our experience has substantiated the premise that comprehensive personnel services can be made available to students living in housing units other than dormitories. Dormitories are not touchstones to adequate personnel procedure. It would be nearer the truth to state that dormitory procedures and personnel formulations are equally in need of drastic re-evaluation. Dormitories simplify a good many problems but the long term reality of students in rooming-houses and private homes must be realistically faced. Until adequate dormitory space is available and within the means of all students, there will remain a student group whose personnel needs are both pressing and important.

# RELATIONSHIP OF THE COLLEGE ADMINISTRATION WITH FRATERNITIES AND SORORITIES

GEORGE S. BEERY

Dean of Students, Drake University

There is a tide in the affairs of men  
Which, taken at its flood  
Leads on to fortune.  
Omitted, all the voyage of their life  
Is bound in shallows and in miseries.  
On such a full sea are we now afloat,  
And we must seize the current when it serves  
Or lose our venture.

TRULY we are afloat on such a full sea of expanding enrollments, increased incomes, and increased opportunities for improving personnel agencies and activities. We are also afloat on a full sea of increased student personnel problems, of administrative responsibilities, of problems of variegated composition of student bodies and of the challenge of establishing and maintaining the high idealism and student morale which we have announced as our mutual goals.

It is not my wish to present a "white paper" for fraternities and sororities nor to go into a long discussion of some factors which may seem to make them advantageous or disadvantageous to campus life. Magazines of national circulation, including *Life* and *Time*, have recently given some publicity to certain aspects of fraternity and sorority life. Most of us are familiar with the opening blast in the anti-sorority publicity which was so definitely pronounced by Mrs. Glen Frank. This anti-Greek letter organization publicity has received the attention of the National Officers of fraternities and sororities and they have diligently prepared their constructive arguments and rebuttals. I shall refer more specifically to some of these pro and con arguments a little later.



That we now have a student body which we never before envisioned is indisputable. For comparison, in pre-war England the largest number of people to be enrolled in colleges and universities was approximately forty thousand or around one to one thousand of population. In the United States this year, according to figures released by the United States Office of Education, we have approximately one in seventy in college and the prediction, based on estimates of the Veterans Administration, that the peak of veterans enrollment will be reached in 1949 and 1950 indicates that in the next few years approximately one in fifty of our general population will be enrolled in college. The figure is all the more impressive when we consider that, of the age group eighteen to twenty-five, approximately one in six is in college.

According to the *World Almanac* of 1947 there are over thirty-six hundred active chapters of social fraternities and sororities with a total membership of over one million three hundred thousand. This does not mean that we have one million three hundred thousand of our present student body of over two million in fraternities (permit me to use the term fraternity as a generic term including both men's and women's organizations). It does mean, however, that there is a potential of this many people to be reckoned with as a rather effectively organized pressure group or body of pressure groups. The present active fraternity membership on college campuses is in the hundreds of thousands and cannot be readily brushed aside.

A recent publication issued by the Inter-Fraternity Research and Advisory Council and entitled *The American College Fraternity—What College Administrators Say Of It* represents an attempt on the part of national fraternities to meet the anti-fraternity publicity to which I have previously alluded. This bulletin quotes verbatim both favorable and unfavorable comments from college administrators in reply to the premise stated by the National Inter-Fraternity Conference in its investigation.

We believe that fraternities constitute an integral part of the life of the colleges. We believe that, in providing rooming,

dining, and social facilities, they are acting as an adjunct to the college, performing a college function. We believe that, while they may and do have many social aspects, nevertheless the social, no less than the literary, the scientific, the philosophical, or the technical, is a part of educational development and a part of the college function. In short, we believe that, in all their activities, including social activities, fraternities participate in and supplement the work of the college in promoting scholarship, leadership, and character.

I want to incorporate a couple of these representative comments in my discussion. For example:

The American College Greek Letter fraternity is one of the oldest student organized activities on the college campus. It has continually attracted the interest and active participation of the finest and ablest students in the student body. It has become an integral part of the total educational program of the college, and its contribution to the education and training of the student has been distinctly wholesome and varied.

We all know that no human institution is without its faults and that individual fraternity chapters do not always measure up to their obligations and opportunities. But by and large, when they fail to do so it is through careless leadership or neglect on the part of the college in which they are located.

Fraternity chapters serve as an excellent medium for the further encouragement and development of scholarship, leadership, and citizenship. They offer the best experience for learning the art of cooperative, harmonious living and for developing the skills basic to better human relations. People are legion who believe that our fundamental problems today lie, in the first analysis, in the area of human relations and that we must learn to live and work and play together harmoniously and effectively if we are to build a better society. In sharing this view, I can think of no better instrument for developing this skill of cooperative living than the fraternity chapter.

The adverse opinions are represented by such statements as:

The difficulty seems to be that while we all agree that fraternities have high ideals and would like to see them live up to these ideals without compulsion, the very independence which they insist upon all too often is accompanied by a reluctance to assume responsibility for the activities of their members.

My experience has convinced me that they too infrequently actually achieve their purposes in their working programs.

Fraternities offer a great possibility for promoting scholarship, leadership, and character. Unfortunately, however, the

fraternities have not all worked out programs that would contribute to these purposes.

Potentially, fraternities are educational in their contribution to the campus. Actually, many of them fall far short of this end, and for that reason they probably are not recognized by either their own members or by the college community at large for their real worth. If the national fraternity organizations could keep a closer touch on their groups so that the objectives of the national officers could be worked out, there is no limit to the constructive good which a fraternity might accomplish in a college setup.

Last week in a Mid-Western newspaper the following item appeared.

Consideration of increasing the number of fraternities at the X University here will be delayed until the present fraternities "show they have grown up," Professor Z, chairman of the committee on student life said Friday.

Professor Z believes the fraternities put themselves on a spot through their foolishness and brutality during "hell week" or initiation ceremonies.

"It seems to me too many fraternity members have an erroneous notion as to how important the fraternity is to the institution," he said. "They live in a smug little existence, not realizing that institutions themselves have changed."

Professor Z said at a meeting of his committee Thursday on the subject of fraternity expansion, there was speculation whether the university ought to have any fraternities, let alone an expansion of the existing ones.

"There is a definite feeling in the committee that the fraternity has outlived its usefulness," he added.

Professor Z's committee has been named to work out proposed changes in requirements for establishment of fraternities in X University, and to investigate the need for additional controls on existing fraternities.

In relation to my basic philosophy relative to relationships between the college fraternities "it couldn't happen here" if the college administration assumed the full responsibility in this area. In a word, as long as the Master's hand is on the helm the ship does not run into the rocks.

Tuesday of this week I received the following statement from the Secretary of the National Inter-Fraternity Conference which is a very excellent expression of the general philosophy which I am attempting to further.

In my professional life I am a College President and have worked as that and as a Dean in colleges in which fraternities were represented and in my dual capacity as a Conference Officer and a College Executive, I would say that the "powers and responsibilities which a college may grant to the individual fraternities and Interfraternity Councils" are only those which the college administration itself feels it wise to grant. In general, the greatest success is achieved by a cooperative endeavor based upon a sound understanding by the administration and fraternities and a full recognition by the fraternities that they exist on university campuses at the will of the educational administration.

In the basic relationships between the college administration and fraternities it is my firm conviction that fraternities and sororities exist only at the behest of, the discretion of, and with the permission of the college administration. It must be granted that students will organize into some form of social group regardless of the name that we give the group, but again such regularly organized groups just do not exist unless the college administration fosters them. This responsibility to the parent campus has been recognized by the American Association of Deans and Advisors of Men and the National Inter-Fraternity Conference as follows:

We consider the fraternity responsible for a positive contribution to the primary functions of the colleges and universities, and therefore under an obligation to encourage the most complete personal development of its members, intellectual, physical and social.

Therefore, we declare:

1. That the objectives and activities of the fraternity should be in entire accord with the aims and purposes of the institutions at which it has chapters.
2. That the primary loyalty and responsibility of a student in his relations with his institution, and that the association of any group of students as a chapter of a fraternity involves the definite responsibility of the group for the conduct of the individual.

Where there is any uncertainty as to where the primary loyalty of the fraternity member is really fixed there seems to be a sort of mutual suspicion between these two enterprises—college administrations and fraternities—who are joined in personal and professional objectives. Representatives of the Association of American Colleges and the National Inter-Fraternity

Conference have drawn up a definition of the reciprocal relations between college and fraternity. The preamble and a couple of sections in this give us further enlightenment on our basic premise.

### Preamble

Formal education, by organizing and concentrating human experience, develops the many desirable potentialities of the student. It stimulates self-expression and an appreciation of all cultural achievement; it fosters self-government, and develops responsible leadership; it encourages the attitudes and imparts the technique needed for self-development, and strengthens zeal for service.

### Article I

#### The Obligation of the College to the Group and Its Members

Sec. 1. Since man is to operate in and with groups, these objectives can be achieved most effectively in organized group life. The college must therefore recognize, as an essential feature of the educational process, a properly organized and maintained student group life. Accordingly, the college should delegate to such groups as large a degree of responsible control of their members as they are capable of exercising with profit to the individual, the group and the college.

Sec. 2. Self-government and individual self-control are essentials of good citizenship, and therefore should be primary objectives of college training. The fraternity chapter, when it realizes fully its potentialities, presents a unique opportunity for discipline in group living and for the practice of responsible self-government. The educational process will be furthered, if the college will utilize and strengthen all resident groups to this end.

### Article II

Sec. 1. Any organized group of undergraduates, irrespective of the purpose for which it is formed, is an integral part of the college.

Last year I spoke before the National Convention of Delta Zeta on the topic "If Privilege Then Responsibility." I challenged the fraternities to make good on their announced standards and ideals. One of the units of the Decalog of Fraternity Policy adopted by the National Inter-Fraternity Conference was called to their attention, a fact which it is our responsibility as college administrators to see brought into actuality. This is it:

- II. The college fraternity must regard itself as an integral part of the institution in which it is located. It not only must be amenable to the rules and regulations of the college institution, but must share in all the college responsibilities of the undergraduate. The college fraternity must match the discipline of the college administration, and must accept the added responsibility incident to the supervision of group life in the chapter house. Furthermore, the college fraternity, with complete loyalty and allegiance to the college which nurtures it, has the duty of supporting in every possible way the institution of which it is a part.

Since we give fraternities privileges, we must demand of them responsibility. Earlier this week I had a fraternity president come in to my office and slump down in a chair as if the weight of the world were on his shoulders. Said he, "I am tired, beastly tired of trying to make men live up to University and fraternity standards and ideals and I won't give up because I know I am right." And then we chatted over some mutual problems. I had had occasion to ask the fraternity president to help in the social adjustment of a boy to whom group approval was perhaps the only disciplining force that he had known. I must confess that I do not have presidents of fraternities and sororities come in every day and express the diligence with which they uphold university regulations. That would be Utopia. However, I insist that there must be a close personal relationship between college administrators and fraternities to the end that the University and its benefits take precedence over all other group objectives. No college can afford to relegate to the area of non-attention the ever growing groups and problems of fraternities or sororities. The University must take advantage of its position of over-all responsibility for the stimulation and direction of the individual lives of its students.

If we find that a particular group becomes obstreperous and needs disciplining the National Officers of the offending group are ready and willing to come in and to apply the necessary pressure and to remedy the situation. In a recent survey of National Officers of fraternities and sororities I was pleased to have National Presidents say these things:

You may be very sure that the fraternity insists to its chapters that their first responsibility is to the college which makes their existence possible. In cases when administrative decisions or regulations conflict with our fraternities' regulations, it is the fraternity and not the college which is expected to make adjustments.

They (college administrations) have a tendency to use the organizations to help them put over their plans and objectives, including housing, but they seldom consider the organizations unless they feel it is primarily to their own advantage. In housing procedures I think it is true. I believe also that it is in the province of the administrations to help control unwise building on the part of local groups and local alumnae. National organizations many times differ with their alumnae but it is not always possible to control them. The national officers have broad and sometimes painful experiences back of them, but actives and local alums do not and they are not always happy with supervision.

However, is it not a bit embarrassing for us to confess that we cannot handle groups on our own campus? It is rather satisfying to note that we do have the support of National Officers and I do not hesitate to confess that I have on a couple of occasions called on them for assistance which was readily given. The University did not lose face and the group was strengthened. Perhaps I should say "refined by fire."

Thus far I have stressed the grave responsibility of the college administration in assuming complete direction of the social groups on its campus. I have not mentioned housing specifically, although I believe a University should agree upon a definite policy for the housing of its fraternities—a policy which will avoid the "mortgage-happy days" of the "roaring twenties" which resulted in the "fraternity bust-up days" of the thirties—a policy which will include fraternity housing as a definite and integral part of university housing, subject to all-university regulations. I have not mentioned leadership training, which we turn over too frequently to the organized groups and which we neglect ourselves. I have not mentioned improvement in scholastic standards, to which the fraternities prescribe and to which we are mutually bound by face-saving and nation-saving obligations. I have not mentioned character training, which is quite a sore spot on many over-expanded campuses. These are all important, and we cannot neglect our duties in their behalf.

What I have tried to present is the rather discomforting fact that the college administration is responsible for a firm and progressive attitude toward fraternities and is also responsible for some of the ills which have resulted by reason of the lack of a firm and progressive policy of action. I have quoted college administrators as well as fraternity officers relative to the relationships between college administrators and fraternities and sororities. I have presented by basic philosophy relative to the attitudes which should be assumed, prerogatives which should obtain, and the responsibilities which are inescapable. We may now well search our souls for a solution in what appeals to me as being a rather potent era of either great good or great evil. My own personal feeling is that we must use these organizations for good, we must hold them to the responsibilities for which we give them privilege and we must ourselves assume a very realistic obligation in all of the areas where we have mutual personal and professed objectives.

Recently I heard a college president speak before an Inter-Fraternity Council dinner. He stated that fraternities have a place and a service to perform, if they will do two things: (1) Instill significance into the life of the college or university. (2) Produce the fraternity men (or women) who are the persons of sanity, balance and tough reasoning power needed in this age. In a world which seems to be filled with insanity, someone has to stay sane.

In keeping with that note and in recognition of the values of the theme of this conference—Developing Human Resources in Democracy—shall we not as college personnel workers and administrators “seize the current” of our obligations, responsibilities, and opportunities for firm guidance and supervision of fraternities and sororities, and save our “venture” of enlightened and useful student resources in democracy?



# THE COLLEGE PERSONNEL WORKER'S RESPONSIBILITY FOR THE IMPROVEMENT OF READING

RUTH STRANG

Professor of Education, Teachers College, Columbia University

ALMOST every week I receive letters like the following from college personnel workers who have been suddenly given responsibility for the improvement of reading. One newly appointed Co-Director writes:

At the present, I am employed by \_\_\_\_\_ College as co-director of guidance. There is one aspect of our guidance program on which I need your advice. We are starting a study skills and remedial reading program here for our students. Could you possibly send me a list of references and materials that would be helpful.

A personnel worker of more experience described his problem as follows:

I am endeavoring to start the first remedial program for college freshmen at our college. Some 80 students are enrolled in the four sections of Remedial Reading and I have administered your Examiner's Reading Diagnostic Record for High School and College. Would you be kind enough to suggest any study that I might do or any new books in the field that I should be familiar with. I find the Diagnostic Record excellent for both teachers and students. Sometimes students come in thinking their reading is good enough, but in filling out the blank they discover their weaknesses. This process of self-appraisal seems to motivate them.

A Dean of Women presented the following problem:

We have a big reading problem on our campus and need a plan for improvement including a reading laboratory. We have no one on the campus to do this except myself. How can I begin developing a sound program?

These letters suggest the three main types of responsibility

college personnel workers have had to assume for the improvement of reading on their campuses:

1. As a consultant on the college-wide reading program.
2. As teacher of special reading groups.
3. As a reading counselor.

*The Personnel Worker as Consultant on the  
College-Wide Reading Program*

The Personnel worker needs to understand the all-college program for the improvement of reading so that he can help the administrator to provide the kind of program that will be the most effective for the institution. The personnel worker brings to an individual conference with the administrator, a policy committee, or a faculty meeting, information of the following kind: One of the aims of college education is growth in reading efficiency and interests; reading development during college years is important because it has personal and social values—it helps the student to understand himself, other persons, and the world in which he lives; reading today has world-wide significance.

Reading supplements and reinforces other avenues of learning—discussion, motion pictures, film strips, radio, excursions. Reading, writing, speaking and listening are interrelated. For example, discussion often shows the individual how inadequate his knowledge of the subject is. Consequently he reads to fill in gaps in his knowledge and to be able to take part more effectively in the next discussion. In reading, the student gets a sense of structure and style; he sees how different authors build paragraphs and convey thoughts. In writing, he reverses the process and creates the pattern of thought himself. It is also well known that a movie increases the demand for the book from which it is made.

Growth is possible no matter what the student's initial reading abilities are. A stimulating college environment itself promotes growth in reading. If the classes, whether dealing with modern problems or with the classics, relate college to life, the student will read with a will to comprehend and to apply what he reads. If the informal student activities lead to discussion

and to projects that appeal to the student as worthwhile, he will of necessity have to read more efficiently in order to find time for all of the things that he wants to do. In the dormitory, a browsing library of significant current books and magazines invites wide reading.

A stimulating environment can be reinforced by instruction and practice in the reading of each subject. Every college teacher should be a teacher of reading in his subject. Every subject has its special reading problems. There is a technical vocabulary to be acquired in the subject. In English the problem of how to recreate characters from the author's clues is paramount to the appreciation of novels, short stories, and biographies. In mathematics and science the reading of equations and formulas requires special study.<sup>1</sup> Obviously the personnel worker will not be responsible for analyzing reading difficulties and for suggesting methods of instruction in each of the content fields. He will, however, help faculty members to realize the importance of teaching reading in their own fields. If he has had some preparation in methods of improving reading, he can go further and work out with each department effective procedures. He will emphasize the importance of students having a definite and vital purpose for reading, and of first-hand experience with new concepts introduced in the text or reference reading. From his file of test results the personnel worker may make charts showing teachers the wide range of reading ability represented in their classes. These facts will impress upon them the need for individualizing reading assignments and for providing books covering a wide range of reading difficulty. With the faculty as a whole, the personnel worker will advocate the integration of reading, curriculum, and instruction.

If the question of how to accomplish this integrated program arises, the personnel worker should be prepared to describe several different types of reading programs. The program at the University of Chicago is outstanding for its recognition of reading, writing, and speaking as an intrinsic part of the college

---

<sup>1</sup> Gray, William S., *Reading in General Education*. Washington: American Council on Education, 1940. pp. 113-185

program—not as an “extra” introduced on a voluntary non-credit basis.

If the college curriculum is organized around the broad fields, the teaching of reading would naturally belong in the language arts area. If the traditional subject matter organization prevails, the best qualified faculty member might take special responsibility for helping all students to improve in reading, recognizing that college-level competency can be attained only by reading in every subject.

### *The Personnel Worker as Teacher of Special Reading Groups*

Faced with students who are not succeeding in their college work because of poor reading, the college administrator frequently turns to the personnel worker for help on this problem. If the personnel worker has had no preparation in the field of reading, he is in a difficult position. He cannot learn the psychology and pedagogy, the materials and procedures of remedial reading in three easy lessons. Yet he is frequently the best qualified person in the college to do the job. So he draws on his background of testing and other methods of understanding students and reads feverishly in the field.

He is first confronted with the problem of who should be admitted to the special groups. To decide this, he will not rigidly apply a single criterion. Instead he will consider each student individually and take into account the following evidences of the need for special help in reading:

Failure to make progress in reading in the regular college program.

Being unable to keep up with the reading required in college classes.

Scoring in the lowest quarter of all college students taking one or more of the standardized reading tests.

Having a reading age of two years or more below his mental age.

Being referred as needing special help in reading by one of the teachers.

Recognizing his own reading difficulty and requesting special help.

The personnel worker must next decide on certain administrative details, the size of the class, the time of meeting, the frequency and length of the periods, credit or no credit. Whenever possible he will schedule the reading group as a regular part of the student's program, limit the number to about twenty students, give academic credit and have two or more meetings a week. If possible, one of the periods each week should be an individual conference. The exigencies of the situation, however, may require quite different conditions. For example, Professor Irving Anderson of the University of Michigan had about five hundred veterans who needed help in reading and enrolled one hundred and fifty of them in one class. In my own group, the number has become too large to permit time for an individual conference with each student once a week. Although there is evidence that improvement is possible even when conditions are not ideal, the personnel worker can try to make conditions as ideal as possible.

After the personnel worker has attended to all of the administrative details, he is faced with the problem of what to do with these seriously retarded readers in the class period. He will not go far afield if he includes the following features.

1. An introductory period in which he emphasizes the following points:
  - a) They are in good company—many persons, some very prominent in business and professions are interested in improving their reading.
  - b) They can improve their reading: evidence from test results and statements made by students in former classes.
  - c) They can gain in efficiency by giving attention to the strategy of study—give suggestions about scheduling time and general study methods.
  - d) The class period will give them practice in getting the author's pattern of thought as they read, and other skills needed to achieve their purposes when reading.
  - e) They will take the initiative and responsibility for studying their reading and making and carrying out plans for improvement keeping a diary record and using the instructor as a consultant and a resource.

2. A standardized test and informal tests to learn more about their reading status and difficulties.
  - a) *Iowa Silent Reading Tests*, New Edition, Advanced Test.  
*Cooperative Reading Comprehension Test C2*.  
*Nelson-Denny Reading Test*.
  - b) Informal group tests in different content fields, read for different purposes.
  - c) Oral reading paragraphs to ascertain difficulties in word recognition, phrasing, comprehension, and attitudes toward reading.<sup>2</sup>
3. Practice in the kind of reading in which they need to improve:
  - a) Group practice.
  - b) Individual practice using files of practice material.
4. Instruction in effective reading methods.
5. Evaluation of progress through a comparable form of initial test, graphs and charts, or written reports and oral discussions of what daily exercises they thought had helped them most.

In conducting classes of this kind, the personnel worker is confronted with the question: What equipment and materials of instruction do I need? With a limited budget, he considers the relative value of the various machines, practice exercises, books, magazines, and pamphlets.

First he may consider a vision screening test. One of the newest of these tests is the *Massachusetts Vision Test*. It takes only a few minutes to give and according to figures supplied by the company, detects large percentages of the individuals who need a more thorough eye examination. The ideal, of course, is to have every applicant for a special reading group have a thorough examination by a competent ophthalmologist.

Second, he is confronted with a number of instruments designed to improve the student's perception of phrases. The *Metron-o-scope* and the Harvard moving picture films, both off the market at present, have been the most widely used

<sup>2</sup> Strang, Ruth, *Examiner's Diagnostic Reading Record for High School and College Students*. New York: Teachers College Bureau of Publications, 1939.

instruments for this purpose. More recently the Root Near Point Tachistoscope patterned after the device used effectively in the Armed Forces for training perception, has been used for the improvement of reading. The theory underlying this instrument is that by stimulating the individual to look more intently, he learns to perceive words more quickly and accurately. Thereby reading and spelling are improved.

The *Buswell Reading Board* is a pressure device for students who have fallen into unnecessarily slow habits of reading. Those who have used it with college students report its effectiveness for individuals.

All of these instruments have value in individual cases. They interest and motivate some students greatly. Their value for groups has not been proved. There is evidence that students improve in groups in which these instruments are used. There is some evidence that the gains made persist or even increase. But whether these gains can be attributed to the instrument or to the other means of improving reading used has not been determined. In the experiment<sup>3</sup> that kept all conditions constant except the use of the instrument there was no evident advantage of the machine (an individual type of metronoscope) over the use of the same practice material without the machine.

What then should the personnel worker recommend with respect to the purchase of these machines? If he has unlimited funds (in most colleges, a hypothetical situation), he may purchase the *Massachusetts Vision Test*, the new form of ophthalmograph, the individual metronoscope that may be put on the market, the quick projection apparatus, and the *Buswell Reading Board*. These instruments motivate certain students. If, however, his funds are limited, he will buy practice exercises closely resembling the reading jobs students have to do in college. He will spend his money for pamphlets, magazines, and books covering the whole range of reading achievement represented in his group. Thus he will have the material necessary for giving students the practice in reading they need.

<sup>3</sup> Westover, Frederick L., *Controlled Eye Movements versus Practice Exercises in Reading*. New York: Bureau of Publications, Teachers College, Columbia University, 1946.

The class period may be spent in different ways. Usually it would include these features:

1. Group practice and instruction in kinds of reading needed by the group as a whole.
2. Practice and instruction needed by sub-groups.
3. Individual practice as needed. For this, files of practice exercises from which individuals can select the kinds they need.
4. Conferences in which the individual can review his plan for the improvement of his reading and obtain further guidance as to his general strategy of reading, the kind of practice and instruction that will be most helpful to him, and the reading program that will contribute most to his personal development and social usefulness.

#### *The Personnel Worker as a Reading Counselor*

Even in the best special reading group there will be a few students who do not make progress. These require the more individualized clinical treatment. With these students the personnel worker, already competent in counseling and psychotherapy, has an advantage over the more narrowly trained reading specialist. To his general counseling competencies, however, the personnel worker needs to add specialized techniques of diagnosing and treating complex reading difficulties. Starting with the problem as the student sees it, he will explore all possible causes of the reading difficulty—physical factors, early developmental experiences that are influencing the individual's present performance, environmental conditions and relationships that are distracting or depressing the individual and emotional and other personality tendencies that prevent the student from concentrating and from putting forth the effort that effective reading demands.

In working with individual reading problems he will do these things:

1. Give the student an idea of the kind of work he will be doing with him individually.
2. Encourage him to take responsibility for making and carrying out plans for his improvement in reading, using the personnel worker as a consultant or resource.



3. Give a standardized silent reading test, an oral reading test, and informal reading tests to find out how he reads, what kinds of errors he makes, what difficulties he encounters and how he meets them, what conditions have given rise to his reading retardation, and what is his attitude toward himself and toward reading.

4. Provide practice and instruction along the lines indicated by the appraisal and diagnosis. It is very important that the reading tasks be appropriate to his present level of reading ability and geared with his real life interests.

5. Make connections with the student's teachers so that they may help him apply immediately in their classes the new insights and skills he has acquired in the individual conferences.

While helping individual students with their reading problems, the personnel worker may also be talking with other students and faculty members to get their ideas of what should be done. These individual conferences might eventuate in a reading workshop, conference, or series of faculty meetings or departmental study group meetings in which faculty members learn more about how to teach reading in their respective content fields.

Obviously, to conduct a program of this kind requires specialized knowledge and skill on the part of the personnel worker. He should be prepared to do this work. If he does it well, he will win the warm approval of faculty members and students.

# SCHOOL MARKS AND SUCCESS IN ENGINEERING

GEORGE A. PIERSON

Dean of Students and Professor of Education, University of Utah

ENGINEERING colleges are now confronted with grave problems of selection and guidance. The total number of students registered in engineering curricula last year was three and one-half times the number enrolled in 1945. In pre-war times sound distribution practices required a continuous review of the criteria used in selecting young people for this exacting profession. Increased enrollments today demand an understanding of the predictions implicit in the process of guiding this host of prospective engineers.

At the University of Utah we have been studying the fundamental relationships that seem to exist between school marks and success in engineering. We hope that our data will supplement the new Pre-Engineering Inventory recently made available by the Graduate Record Office.

Our investigation was undertaken to discover: (1) the combination of marks earned in high school and the first year of college that makes possible the most efficient prediction of scholastic success in engineering curricula; (2) some indication of the reliability and validity of scholastic achievement in engineering as a criterion of success in the profession; (3) the differences between the marks earned by engineering-school graduates and drop-outs.

In order to arrive at an adequate solution to these problems, the academic records were assembled for the 463 pre-war graduates of the School of Engineering of the University of Utah for the ten-year period 1932 to 1941. All work taken in college was used to compute a grade-point ratio for each engineer and this ratio was employed throughout the study as an index of general scholarship.

Standard statistical techniques were then used to describe the high-school achievement of the typical engineer and to discover the relationship between marks earned in high-school subjects and general scholarship in Engineering. The same standard procedures were also used to determine the relationship between marks awarded during the first year of college and the engineering grade-point ratio.

In addition, a judgment of the postgraduation success was secured for 320 of the 463 engineers studied. Each man was rated, in terms of a five-point scale, by the faculty member best qualified to evaluate his particular accomplishments. These success ratings were then correlated with general scholarship in engineering and with the grade-point ratios achieved in high school, at the end of the first quarter, and at the close of the freshman year.

The final educational status was also determined for the 1,237 students enrolled in the prescribed first-quarter courses in engineering from 1928 to 1937. Complete academic records were assembled for 337 of these students who dropped out of engineering courses. Standard statistical techniques were also employed to compare the scholastic achievement of graduates and drop-outs.

The following is a brief resume of our findings:

1. The graduates of the School of Engineering of the University of Utah included in this investigation proved to be high-school students of superior achievement. Five-sixths of the engineers studied equaled or exceeded a grade-point ratio in high school of 1.6 (B -). The average 1932 to 1941 graduate achieved a high-school ratio of 2.1 (B +).

2. The grade-point ratio achieved in high school was discovered to be more closely related to general scholarship in Engineering than the marks earned in any particular high-school subject. Not only that, but general scholarship in Engineering could be predicted as efficiently from the high-school grade-point ratio as from marks earned in English, Mathematics, Physics, and Chemistry combined in a multiple-regression equation. The zero-order coefficient of correlation proved to be .58 and the coefficient of multiple correlation was found to be only .52.

3. Marks earned in high-school English were discovered to be as closely related to general achievement in Engineering as marks earned in Mathematics or the sciences. High-school marks in Mechanical Drawing and Mechanic Arts were found to have little or no relationship to the engineering grade-point ratio.

4. The grade-point ratio for the typical engineer fell from 2.1 in high school to 1.55 the first quarter in college. During the first quarter in residence, one-fifth of the graduates studied failed to achieve a grade-point ratio of 1.0 (C).

5. The first quarter grade-point ratio was more closely related to general scholarship in Engineering than the marks earned in any particular subject; and, for all practical purposes, general scholarship in Engineering could be predicted as efficiently from the first quarter grade-point ratio as from the first quarter marks in Chemistry, English Composition, Mathematics, and Engineering Drawing combined in a regression equation. The coefficient of correlation between the first quarter ratio and general achievement in Engineering was discovered to be .67 and the coefficient of multiple correlation .74. The difference is insignificant statistically.

6. The first quarter mark in English Composition was discovered to be as closely related to the engineering grade-point ratio as the first quarter mark in Chemistry or Mathematics. Not only that, but the first quarter mark in English proved to be more closely related to general achievement in Engineering than the first quarter mark in Engineering Drawing.

7. Academic adjustment difficulties characterized the first-year performance of the engineers studied. At the end of the freshman year, the mean grade-point ratio remained at 1.4, and one graduate out of every five failed to achieve a "C" average.

8. The first year grade-point ratio was also more closely related to general achievement in Engineering than the ratios earned in any particular subject; and this over-all ratio proved to be as efficient in predicting scholastic achievement in Engineering as the first-year ratios in Mathematics, Chemistry, Engineering Drawing, and English Composition combined in a

regression equation. The zero-order coefficient of correlation proved to be .75 and the coefficient of multiple correlation, .78.

9. General scholarship in Engineering could be predicted almost as efficiently from the grade-point ratio at the end of the first quarter as from the grade-point ratio at the close of the first year. The coefficients of correlation were discovered to be .67 and .75, respectively. The difference is statistically significant.

10. The grade-point ratio in Engineering seemed to be a reasonably reliable measure of scholastic success. When the coefficients of correlation between first-year ratios and general achievement in Engineering were used as measures of reliability, the corresponding indices of reliability were found to compare favorably with the reported reliability coefficients of some of the more widely used standardized tests. The indices of reliability for the grade-point ratio in Engineering ranged from .82 to .88.

11. In addition, the engineering grade-point ratio appeared to be a rather valid basis for predicting success in the practice of engineering. A coefficient of correlation of .43 was obtained between the engineering grade-point ratio and a rating of success in professional practice. This coefficient is as high as the typical coefficient obtained for the relationship between intelligence scores and marks earned in college. While this ratio is limited as an instrument for predicting the success of a particular individual, the follow-up study did reveal that a graduate with a grade-point ratio of less than 1.8 is fourteen times more likely to be rated in practice as "unsuccessful" or as "moderately successful" than an engineer with a higher ratio.

12. The study of drop-outs revealed that the majority of students enrolled from 1928 to 1937 possessed the ability to succeed academically. Seventy-five per cent achieved grade-point ratios of at least 1.0 before graduation, transfer, or drop-out.

13. The student's interest in engineering courses seemed to be an extremely important factor in determining survival. One-fourth of those who began the prescribed first quarter courses in Engineering transferred or dropped out later in good academic standing.

14. Junior-college transfers to the School of Engineering attained higher grade-point ratios at the time of graduation than engineers in general.

15. Even though drop-outs as well as graduates proved to be high-school pupils of superior achievement, the differences between the two groups were always statistically significant. More important, however, was the discovery of a large area of overlapping. Twenty per cent of the students who dropped out of Engineering with grade-point ratios below 1.0 achieved high-school ratios as high as the average for engineering school graduates, or higher. On the other hand, 9 per cent of the graduates earned grade-point ratios in high school below the average for these "failing" drop-outs.

16. The study of drop-outs also indicated that first quarter college marks were much more efficient in predicting future success in Engineering than the marks earned in high school. Differences between the first quarter achievement of drop-outs and graduates were always highly significant and the area of overlapping was greatly reduced. Only 5 per cent of the "failing" drop-outs equaled the average first-quarter ratio attained by engineering school graduates.

17. By the end of the freshman year the scholastic differences between graduates and "failing" drop-outs were found to be clearly established. At that time only one drop-out in a hundred equaled or exceeded the average grade-point ratio attained by the graduates in engineering.

Our investigation seems to imply that marks should be thought of as instruments to be improved and not as outworn devices to be ignored. In spite of the general confusion regarding the place of marks in the teaching-learning process today, we arrived at the following conclusions:

1. The simple grade-point ratio is an index of considerable importance in selecting and guiding prospective engineers. In high school it is the best single predictor of success in college courses in Engineering. It is superior to the marks earned in any particular high-school subject including Science or Mathematics. During the first year of college it is more closely related to general scholarship in Engineering than the marks

earned in Mathematics, Chemistry, English, or Engineering Drawing. The grade-point ratio achieved at the time of graduation from one of the engineering curricula is as indicative of success in professional practice as intelligence test scores are indicative of scholastic achievement in college. This simple ratio is just as efficient, moreover, in predicting general scholarship in Engineering as a more complicated multiple-regression equation which combines mathematically marks earned in high school or marks earned during the first year of college. As a rule, therefore, students should be guided in terms of their general scholastic achievement rather than in terms of the marks earned in any particular high-school or college subject.

2. Even though the high-school grade-point ratio is the best single indicator of a student's probable success in college courses in Engineering, it is not comprehensive enough to be used as the only selective criterion. This index of general scholarship should be supported by scores made on a battery of achievement and intelligence tests, such as the *Pre-Engineering Inventory*, and by some measure of a student's pattern of interests.

3. The multiplicity of factors which should be taken into account in selecting students for engineering training re-emphasizes the need for guidance and counseling services during the last year of high school and the first year of college. The findings imply that the high-school grade-point ratio and various test scores should be thought of as devices for improving the quality of the group selected. These criteria may have little value in predicting the success in engineering of a given individual. The selection of engineering as a vocational objective still remains a kind of "clinical" problem, that is, a "judgment" based upon all of the data available for a particular individual.

# THE GRADUATE RECORD EXAMINATION VS. OTHER MEASURES OF APTITUDE AND ACHIEVEMENT

JOSEPH C. HESTON

Director, Bureau of Testing and Research, DePauw University

## *Introduction*

THE necessity for any detailed introductory material in this paper has largely been dissipated by the excellent address delivered at our luncheon session yesterday by Dr. Vaughn on the general topic of the *Graduate Record Examination*. This has laid the background for this present paper and has thereby greatly simplified my task in discussing this experimental set-up.

In May, 1946, fifty-two colleges and universities, widely distributed over the nation, cooperated in the Inquiry into Post-war Conditions in American Colleges by administering to all of their graduating seniors and second-semester sophomores the new form of the *Graduate Record Examination* now designated as the *Tests of General Education*. Many of our members here today have participated in the administration and research evaluation of the Inquiry program. It is, therefore, with a feeling that I am bringing pertinent material to this group that I should like to present some of the data available from our own institutional analysis of this Inquiry program. We shall not touch upon some of the larger general problems inherent in the Inquiry, such as the demonstrated relationship between senior and sophomore performance on the examinations, the relationship between the performance of the men and women students, nor the profiles for the subject matter major groups that might legitimately be constructed. Rather, we are here concerned with relationships between the *Graduate Record Examinations* and other well-recognized tests of ability and school achieve-



ment. Further, we shall study the relationship between these various tests and the so-called criterion of academic achievement, the scholastic grade-average. As a consequence of this approach, much of the data in this paper are of a correlational nature and since correlations are confusing when presented en masse, the writer has prepared several mimeographed tables which bear in essence the results of this analysis.

### *The Data Available for This Analysis*

At DePauw University in May, 1946, we administered the Inquiry program of the *Graduate Record Examination Tests of General Education* to a group of 138 seniors and 280 sophomores. Inasmuch as the great majority of the students in these classifications were women, many of our larger group analyses presented below will be restricted to women's scores. This we think is justified because of the fact, established both at DePauw and throughout the nation, of a decided differential in the performance between men and women students on these tests.

For the sophomore students involved in this DePauw program we have available a large number of entrance test scores, specifically the *A.C.E. Psychological Examination*, the complete *Cooperative English Test*, and the three *Cooperative Tests of General Achievement*, in Social Studies, Natural Science, and Mathematics. These, along with other tests not pertinent to this investigation, were administered to the students upon entrance to the University as freshmen. A composite figure representing a student's performance on the whole battery is available in the form of a variable we have labeled the "Ability Index." This is composed of the T-score (local) earned on the *A.C.E. Psychological Examination*, plus twice the T-score for the combined total on all of the four Achievement tests. T-scores, based on a mean of 50 are used, so the expected mean is 150 for the typically average student. The weighting has been entirely arbitrary, but does have a fair degree of validity in actual practice.

For the senior students we do not have the results from this complete battery of tests. We could have secured for these

senior students scores on the *A.C.E. Psychological Examination*, which may have been taken at different times and therefore involved different forms of the examinations. As the data presented below, and other data collected in our office, reveal the A.C.E. is less significant than the achievement tests, we have consequently not utilized these data in the present analysis for seniors.

The other chief category of data available for this analysis is the familiar and much abused scholastic grade-average. We frankly recognize at the outset the potential unreliability of this device as a criterion of true academic achievement. We agree perfectly with Williamson (*How to Counsel Students*, page 25) when he says "When the guesses of different instructors are combined to derive an average (hour-point ratio, grade-point average, etc.) we have much the same result as though we added apples and pears, multiplied the sum by doorknobs, and divided by peanuts. The result is arithmetical truth but psychological nonsense." This condition is deplorable and we have elsewhere conducted research to demonstrate certain reasons for the inconsistency of grade averages. However, in the present instance there is no *additional* or alternative criterion which might be utilized to check the validity of both our entrance examinations and the Graduate Record Tests. We are, therefore, by default forced to accept this scholastic average as the only available criterion of educational development. In computing this average, P.H.R. (point-hour-ratio), an A = 3 points, B = 2 points, C = 1 point, D = zero points, and F = -2 points.

We selected three large basic groups of women liberal arts students for the major portion of this analysis. At the sophomore level we have two groups, each composed of 100 women, and at the senior level one group composed of 100 women. In each case the selection of these women students has been such as to ensure a sample as representative as possible of the available group of students on the General Educational Index, the composite figure representing overall achievement on the Tests of General Education. The data presented in Table 1 describe statistically the characteristics of these groups. It will be seen

TABLE 1  
*Basic Descriptive Data for the Principal Groups Studied*  
*(DePauw University Women)*

Group	(N)	Means			S.D.		
		G.R.E. Index	Ability Index	P.H.R. (Grades)	G.R.E. Index	Ability Index	P.H.R.
I Sophomore Women	(100)	466.2	154.4	1.50	86.8	27.6	.54
II Sophomore Women	(100)	457.8	151.8	1.39	74.2	25.3	.56
Combined I + II	(200)	462.0	153.0	1.46	80.8	26.5	.55
III Senior Women	(100)	497.9	....	1.76	93.8	.	.45

that the two sophomore groups are very comparable to each other. The two sophomore groups of 100 each have been compared with each other wherever possible to demonstrate the degree of consistency obtaining in the analyses. Both the sophomore and senior groups correspond fairly closely to the means and standard deviations in the women's norms established by the nation-wide Graduate Record Inquiry program.

### *Test Interrelations*

In this section we shall study the intercorrelations existing between the various tests on the *Graduate Record Examination* and the Cooperative entrance test battery. In Table 2 we have the correlations obtained between the G.R.E. Index and other indices secured from the entrance test battery. The correlations are consistent between Group I and Group II, hence one is justified in taking the combined correlations as indicative of the overall relationship between the indices. The coefficient of .87 between the G.R.E. Index and the Ability Index is almost high enough to be a satisfactory reliability coefficient for a

TABLE 2  
*The G.R.E. General Educational Index vs. Entrance Test Indices*  
*(Based on DePauw University Sophomore Women)*

Variables Correlated	Correlation Coefficients		
	Group I (N = 100)	Group II (N = 100)	Combined I + II
GRE Index vs. Ability Index . . . . .	.88	.86	.87
GRE Index vs. ACE Psych. Exam. . . . .	.77	.79	.78
GRE Index vs. 3 Achievement Tests . . .	.86	.82	.84

single test. When one recalls these two index scores were earned nearly two years apart, it is apparent that the two indices thus obtained are very highly related. The Ability Index may be broken down into its components, the *A.C.E. Psychological Examination*, and the *Cooperative Achievement Tests*. It appears the G.R.E. Index is slightly more related to the index obtained from achievement tests ( $r = .84$ ) than to the *A.C.E. Psychological Examination* ( $r = .78$ ). This higher degree of relationship between the *Cooperative Achievement Tests* and the Graduate Record Index is understood when one recalls that in both sets of tests much of the score depends upon one's ability not only to recall knowledge, but to read, comprehend, and interpret questions based on material supplied in

TABLE 3  
*Intercorrelations Among the G.R.E. Tests of General Education*  
(Based on Group I: 100 DePauw University Sophomore Women)

	Math.	P.S.	B.S.	Soc.	Lit.	Arts	Exp.	Voc.
Mathematics .....	..	.55	.44	.51	.36	.35	.52	.38
Physical Science .....	.55	..	.49	.43	.20	.40	.32	.29
Biological Science .....	.44	.49	..	.57	.42	.42	.46	.50
Social Studies .....	.51	.43	.57	..	.54	.40	.61	.59
Literature .....	.36	.20	.42	.54	..	.39	.53	.54
Arts .....	.35	.40	.42	.40	.39	..	.42	.52
Effective Expression ..	.52	.32	.46	.61	.53	.42	..	.66
Vocabulary .....	.38	.29	.50	.59	.54	.52	.66	..

the tests. Inasmuch as the A.C.E. calls for less of this capacity, its correlation is somewhat lower with the G.R.E. Index.

Table 3 presents the intercorrelations obtained among the G.R.E. tests, as based on Group I (100 Sophomore women). All of the correlations are positive and many of them are of significant proportions. This suggests some sort of a common ability factor operative in all of the tests. This hypothesis will be evaluated and further clarified in a later section of this report, where the correlational matrix of Table 3 is subjected to factor analysis.

Inasmuch as the G.R.E. Tests of General Education and the Cooperative battery of entrance tests have much similarity in their specific test titles, it would seem advisable to ascertain what correlation may exist between tests with parallel titles.

TABLE 4

*Correlations between Various G.R.E. Tests of General Education and  
Corresponding Cooperative Tests  
(Based on DePauw University Sophomore Women)*

Tests Correlated	Correlation Coefficients		
	Group I (N=100)	Group II (N=100)	Combined I+II
GRE Math. vs. Coop. Math. ....	.72	.59	.66
GRE Phys. Sci. vs. Coop. Nat. Sci. ....	.68	.54	.61
GRE Biol. Sci. vs. Coop. Nat. Sci. ....	.58	.52	.55
GRE Soc. Stud. vs. Coop. Soc. Stud. ....	.65	.66	.65
GRE Eff. Exp. vs. Coop. Eng. Mech. ...	.70	.58	.65
GRE Eff. Exp. vs. Coop. Eff. Exp. ....	.69	.55	.63
GRE Vocab. vs. Coop. Vocab. ....	.82	.80	.81
GRE Vocab. vs. ACE Psych. ....	.66	.72	.69

We have, therefore, in Table 4 presented these correlations. The relationship is presented for the separate Groups I and II, and then for all 200 cases combined. Some of the tests are definitely more consistent than others in the correlations demonstrated between the two sub-groups. The most consistent tests are those involving social studies and vocabulary. They are decidedly more consistent than the so-called "objective" concepts represented in the fields of mathematics and science. The highest inter-relationship obtained is between the two vocabulary tests, due undoubtedly to the specificity of content. In all the other tests *interpretation* of reading matter is a factor to be considered.

The corresponding specific tests of the two batteries do not correlate as highly as one might hope for tests involving such similarity of titles. However, if one bears in mind the nearly two-year time lapse occurring between the administration of the two sets of tests, then the obtained correlations are more satisfactory than they would appear to be at first glance. As would be anticipated, the correlations between *specific* corresponding tests in Table 4 are not as high as those obtained between *composite* test indices in Table 2.

#### *Relation between Test Scores and Scholastic Achievement*

The Cooperative entrance tests purport to measure both previous school achievement (at the high-school level) and

future proficiency, i.e., ability to succeed in college work. Likewise the G.R.E. tests are designed to measure both past achievement and future proficiency. Thus it is desirable to establish the relationship between these various tests and scholastic achievement, here indicated by grade averages earned in DePauw University. As stated earlier in this report, the writer holds no brief for the infallibility of grade averages. They are, unfortunately, the only other indication of college success for the groups being currently analyzed.

TABLE 5  
*Validity of General Indices and G.R.E. Tests of General Education in  
Prediction of Cumulated Grade Averages (P.H.R.)  
(Based on Depauw University Women)*

Predictive Variable	Correlation with P H R.			
	Group I (100 Sophs.)	Group II (100 Sophs.)	Combined I + II	Group III (100 Seniors)
(General Indices)				
GRE Index .....	.60	.66	.63	.68
Ability Index .....	.48	.65	.56	..
ACE Psych. Exam. . .	.39	.55	.47	..
3 Achievement Tests .	.48	.66	.56	..
(Specific Tests)				
GRE Math. ....	.45	.48	.46	.41
GRE Phys. Sci. ....	.32	.35	.34	.32
GRE Biol. Sci. ....	.40	.38	.39	.54
GRE Soc. Stud. . . .	.54	.56	.54	.58
GRE Lit. ....	.44	.46	.45	.49
GRE Arts ....	.42	.42	.42	.45
GRE Eff. Exp. ....	.49	.50	.49	.62
GRE Vocab. ....	.45	.54	.49	.59

Table 5 has been divided into two sub-sections. In the top section we present the validity of the various composite test indices as predictors of scholastic success for the two sophomore sub-groups and the group of senior women. In all cases the grade averages used here are *cumulated* averages for the full two years for the sophomores and the full four years of the seniors. Of the four general indices shown in Table 5 it is obvious the G.R.E. Index is not only most highly correlated with grade average, but it is also the most consistent. For some reason which the writer as yet has been unable to ferret out, all of the indices have higher validity coefficients for sophomore Group II than for Group I. Even here, however, the G.R.E.

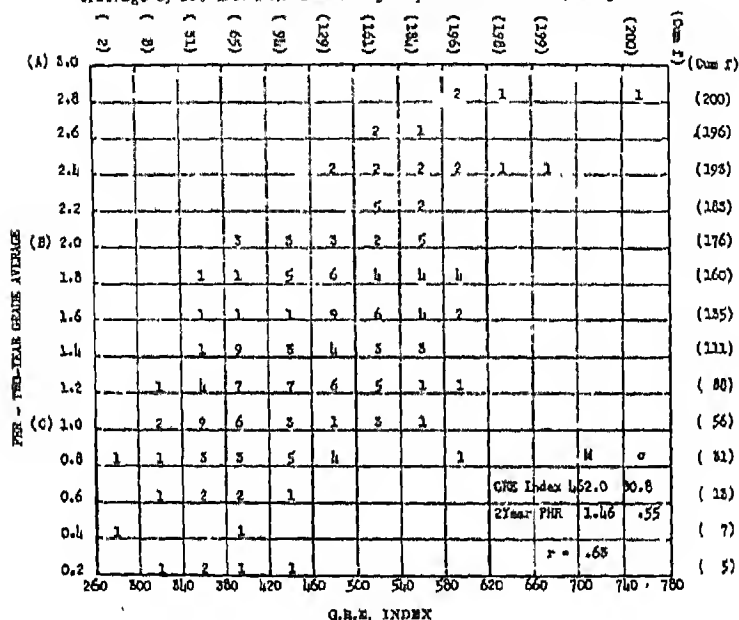
Index shows less variation between Group I and Group II than do the other indices. For Group II there is really no choice between the G.R.E. Index, the Ability Index, and the three Cooperative Achievement Tests as predictors of scholastic success. However, for Group I the G.R.E. Index is definitely superior, and thus, due to its greater consistency between the two groups, it shows a higher relationship for the combined prediction. For the senior women the G.R.E. Index is the only one available. The correlation between the *four*-year grade average and G.R.E. Index is somewhat higher here than for the two-year average of the sophomore women.

In the lower section of Table 5 we present the correlation between each of the eight specific G.R.E. Tests of General Education and the cumulated grade average for each group being analyzed. The consistency between the correlations for each test in the two sophomore sub-groups is noteworthy. The similarity between the validity of each test for sophomore-grade averages and senior-grade averages is also marked. Social Studies, Effectiveness of Expression, and Vocabulary are the three specific tests showing the most relationship with over-all grade averages at DePauw for these women students. The test most obviously inferior in its ability to predict grade averages is the G.R.E. Physical Science Test. This is probably not a deficiency in the test. It is probably an artifact produced by the requirement system in operation at DePauw University. A student is required to elect only one laboratory science. Most of the women students elect a biological science. Consequently few of our population being studied have taken any physical science during their stay at DePauw University. Thus the obtained correlation between G.R.E. Physical Science and DePauw grades is due fundamentally to general ability.

Correlation coefficients are convenient statistical symbols for use in expressing relationship between variables; however they do not give a complete picture of the actual variability which may exist between the two factors being associated. We have, therefore, made use of scatter diagrams in an attempt to depict two of the more important relationships thus far presented. In Figure I we show the relationship for 200 sopho-

more women between the Graduate Record Index and the two-year grade average. The amount of scatter is rather large, although the linearity of the relationship is evident. In Figure II we have presented the analogous situation showing the relationship between the G.R.E. Index and the Ability Index earned on the entrance tests. The correlation here is considerably

Scatter Diagram Showing G.R.E. General Educational Index vs. Two-Year Grade Average of 200 DePauw University Sophomore Women, May 1946



higher, the linearity more pronounced, and the scatter much reduced.

An interesting point for speculation here is the relationship between high grades and success on the *Graduate Record Examination Index*. If these women were graduating with their present grade averages we would award 20 per cent of them a diploma with "distinction," a 2.0 average. Six of these with "distinction" averages actually earned a G.R.E. Index below the national average. A similar analysis of our entire senior class last year produced a parallel result. Two brief hypotheses



may be offered in explanation. One is that the G.R.E. tests depend more upon native intelligence or capacity than do the cumulated grade averages, because in the latter such factors as motivation and personality are known to operate. The other hypothesis is that in some departmental areas high grades are given with a greater degree of regularity than in others. Conse-

*Scatter Diagram Showing G.R.E. General Educational Index vs. Ability Index on Freshman Entrance Tests of 200 DePauw University Sophomore Women, May 1946*

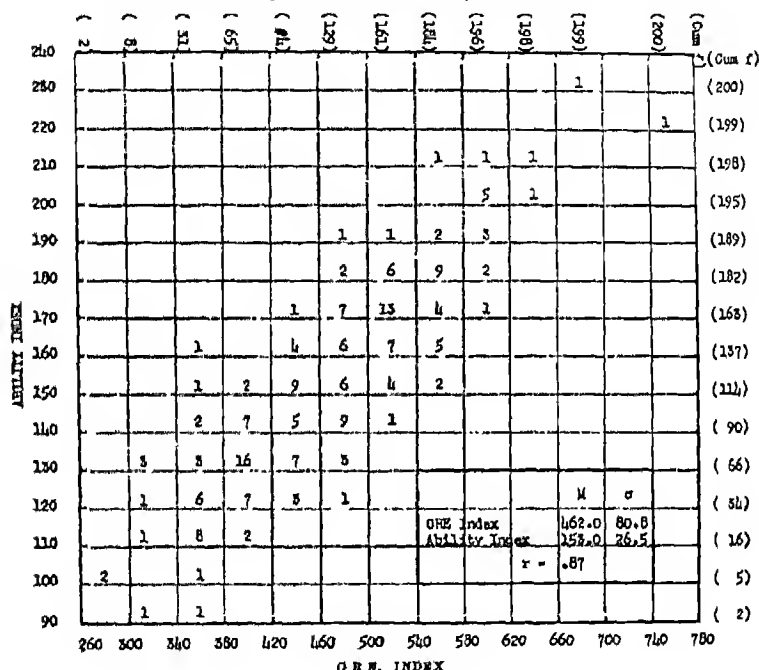


TABLE 6

*Validity of G.R.E. Tests of General Education and Cooperative Tests  
in Prediction of Specific Course Grades  
(Based on DePauw University Sophomore Men and Women)*

Specific Course	(N)	Sem. Hrs.	Correlation with Grades	
			r G.R.E. Test	r Coop. Test
English Composition (99)		6	.56 Eff. Expression	.45 Eff. Expression
English Composition (99)		6	.....	.52 Eng. Mechanics
English Writers (64)		6	.57 Literature	.....
Fundamentals of Art (56)		3	20 Arts	.....
Principles of Economics (22)		6	.65 Soc. Stud.	.57 Soc. Stud.
American Government (46)		6	.76 Soc. Stud.	.62 Soc. Stud.
European History (92)		6	.52 Soc. Stud.	.58 Soc. Stud.
Introductory Sociology (87)		6	.56 Soc. Stud.	.63 Soc. Stud.
Introductory Psychology (79)		4	.37 Biol. Sci.	.40 Nat. Sci.
Introductory Psychology (79)		4	.58 Soc. Stud.	.59 Soc. Stud.
General Chemistry (34)		9-10	.60 Phys. Sci.	.48 Nat. Sci.
General Botany (92)		8	.37 Biol. Sci.	.35 Nat. Sci.
General Zoology (86)		8	.56 Biol. Sci.	.61 Nat. Sci.
College Algebra (Sem. I) (25)		2	.45 Math.	.27 Math.

cessful than the Cooperative test in its correlation with grade averages. In Table 6, due to the smaller number of available subjects in each group, both men and women students have been utilized.

#### *Factors in the G.R.E. Tests*

A final point of interest is the investigation of the extent to which general ability plays a part in the capacity of the G.R.E. tests to predict academic success. As a minor approach to this problem we have computed partial correlations for two of the correlations appearing in Table 6. In this approach it has been assumed that general capacity could be held constant by introducing into the partial correlation the relationship of the Ability Index to the other two variables in question. In the first group in Table 6, English Composition, with Ability Index held constant, the original correlation of .56 between English Composition grade and G.R.E. Effectiveness of Expression is reduced to .38.

A parallel analysis for General Zoology, shown near the bottom of Table 6, shows that with the Ability Index held constant, the original correlation of .56 reduces to a partial correlation of .28. In these two cases, English and Zoology, one can assume that a large measure of the original correlation between the

G.R.E. test and the course grade is due to the common factor of general ability.

For a more exhaustive evaluation of this thesis, that a common ability factor exists in all of the eight G.R.E. tests, the writer has subjected the correlational matrix of Table 3 to a factor analysis by the multiple group method described by Thurstone (in *Psychometrika*, June, 1945). This analysis has been presented in Table 7, where we find three orthogonal primary factors. It is apparent that even through the use of this method, which seeks to establish *separate* factors, that factor I

TABLE 7

*Orthogonal Factor Matrix "F" (Rotated) of Primary Traits in G.R.E. Tests  
(Centroid Factor Analysis of Table 3)*

G.R.E. Test	Factor Loadings			$h^2$
	I	II	III	
Mathematics .....	.577	.427	-.130	.532
Physical Science .. .	.404	.659	.128	.614
Biological Science .....	.636	.331	.002	.514
Social Studies .....	.766	.190	-.162	.649
Literature .....	.701	-.109	.035	.505
Arts .....	.564	.181	.410	.519
Effective Expression .....	.802	.001	-.080	.650
Vocabulary .....	.799	-.087	.207	.689
Mean $h^2$ .....	.447	.102	.035	.584

is common to all eight tests. Factor I, having a mean communality of .447, thus accounts for 45 per cent of the variance of the entire battery. It can be identified by examination of the four highest factor loadings, Effectiveness of Expression, Vocabulary, Social Studies, and Literature. One can assume that this is largely a factor involving verbal facility, the capacity to recognize and to comprehend subject matter primarily involving word usage, as contrasted to the use of formulae, equations, and diagrams. Factor I might be regarded tentatively as general "reading comprehension."

Factor II, which accounts for 10 per cent of the variance, is definitely a science-mathematics factor, the highest loading being on Physical Science, with Mathematics second and Biological Science third. Since the Physical Science test has the

highest loading on this factor it must involve something other than mere mathematical comprehension. The writer is not prepared to offer a complete description of this factor at this time.

Factor III has a significant loading on only one variable, the Arts test, hence it is fundamentally measuring a unique trait. Vocabulary has the only other appreciable loading. The entire factor, since it accounts for only  $3\frac{1}{2}$  per cent of the variance, is negligible in its ability to explain the intercorrelation among the tests.

The thesis postulated in the descriptive *Manual* for the *Graduate Record Examination* that Vocabulary is perhaps the best measure of overall capacity is borne out to some extent by this analysis. Vocabulary has a higher communality from these three factors than any other single test in the battery. Our earlier observation that Effectiveness of Expression and Social Studies are the other best predictors of academic success probably hinges on the fact that they have the next highest communalities. The average communality of .584 means these three factors explain approximately 60 per cent of the variance in the correlational matrix.

In conclusion, the writer has not prepared a summary of this paper. He feels rather that the specific data now in your hands in the mimeographed tables and figures are more pertinent individually than when collected into a generalized summary. Two limitations should be kept in mind in evaluating these data. First, the number of cases involved in all of the analyses has been lower than one would like to have for conclusive evidence on the issues raised. For that reason these observations should be repeated, not only at DePauw University, but elsewhere. Secondly, all of the conclusions presented in this paper are subject to the restriction that in every table, except Table 6, the data have been based on women only. Due to demonstrated differences between women and men on the *Graduate Record Examination*, it is necessary that future consideration of these issues should be based on similar analyses where a sufficient number of men is available to make a parallel investigation.

# TRENDS AND DEVELOPMENT OF THE VOCATIONAL AND OTHER INTERESTS OF VETERANS AT WASHINGTON UNIVERSITY

NATHAN KOHN, JR.

Veterans' Counselor, Washington University

THE observations reported were made by the Veterans' Counselor of Washington University, St. Louis, Missouri. The Counseling office has jurisdiction over all of the Veterans on the campus. It is not related to the Veterans Administration in any official way. Veterans refer themselves on a voluntary basis, although certain survey questionnaires to be discussed were given to selected groups of the Veteran population on the campus. Somewhat more than half of the Veterans on the campus have been seen for one reason or another.

This report includes a brief analysis of the Veterans who entered the University prior to February, 1946, and an intensive study of those that have entered since that time. In order to have some criteria to understand their interests and their significance, a brief description of the backgrounds of these Veterans will be given as an introduction to the analysis of their interests.

A background analysis of this group includes such things as parental education, the student's age, health, and prior education, as well as his prior vocational and Army experiences.

Of the Veterans on the campus, both parents of 3 per cent of them were dead, while in something less than three out of four of the homes, both parents were still alive. Of this population, four out of five of the fathers and nine out of ten of the mothers were living.

A majority of the parents of the group were born in the Middle West with more than one out of two born in Missouri, and another one out of four born elsewhere in the Middle West.

Approximately two out of five of both parents did not go beyond the elementary school, while another two out of five had some high-school education. Slightly more than one out of seven of the parents ever attended college. One in twenty of the parents had technical training. This means that of the parents of these boys in college, less than one out of five of the fathers and one out of ten of the mothers had any college experience. More than one-third of the fathers, and just less than half of the mothers had no high-school education.

The age of the Veterans ranged from eighteen to forty-five years with a mean near twenty-two and one-half.

Of the Veteran students choosing Engineering, three out of four had permanent residence in the St. Louis area but of those choosing non-engineering, one out of two had permanent residence there.

Just less than one out of two of this population were attending school before they were inducted into the army. One interpretation of this is that one-half of the men of the group might not have come to college without the assistance of the G.I. Bill of Rights. On the average, the non-engineering group tended to have been out of school the longest, while those who had made no vocational choice had been out of school for significantly less time than the rest of the group. About one out of ten of the whole group had some training other than college before going into service. Interestingly and perhaps significantly, over two out of three of these men had some schooling or training while in the service, and more than one out of four of the total group received college training. Almost two out of five of the men who received training in service were given training in technical skills.

More than 80 per cent of these men reported having held a full-time job. Three out of four of the men reported that they had worked before entering the service, and of these one-third had attempted to go to school at the same time. About one in four of the group was satisfied with his job before the war, and another group of equal size were satisfied with the general area of their pre-war employment. It is interesting that the group who had failed to make a vocational choice tended to feel most satisfied with their previous job.

Questionnaires concerning whether or not the Veteran is attempting to hold a full-time job while going to college were answered ambiguously. Regarding part-time employment, approximately 40 per cent of the men desired it. Interestingly, one out of two of these wanted to work because of family needs, while the rest of them wanted to work in order to have work experience, extra money for recreation, etc., and some because they did not feel that school was a full-time job.

This total group averaged just more than three years in service. Twenty per cent of the group made no definite vocational choice. One out of four of this group had worked in unskilled positions both in the service and in their previous employment, compared to the other four-fifths of the population in which an incidence of one in ten indicated that they had been in skilled labor both in the service and in their previous employment.

More than four out of five of these men made a vocational choice but just less than 60 per cent felt any certainty about this choice. The incidence of insecurity was significantly higher among those who had chosen Engineering. Of the three-fifths who had a fairly definite vocational choice, less than two out of three felt that they had worked out an adequate plan for preparation for their vocation. Less than one out of ten of the group felt that they had or knew how to get adequate vocational information about the various fields of their interests.

Parents, friends, and relatives had made suggestions as to careers, with varying degrees of pressure, to more than 60 per cent of the men. On the whole, they appreciated the interest and felt the suggestions to be helpful. The exception was the group which had as yet made no vocational choice. Four out of five of those making no vocational choice indicated that family or friends had made suggestions as to their vocational choices. In some instances these Veterans felt that they were being coerced into certain fields. Interestingly, they rejected the offered guidance emphatically.

Considered on a five-point scale from professional to non-skilled, nine out of ten of this whole group felt that they were capable of entering a professional field and of being established

in their profession within ten years. This notwithstanding that less than one out of ten of these men had any advance credits beyond the Freshman year in college.

Less than one-fifth of the group felt that programs and activities to meet other students and participate with them would not be of value, and three out of five felt that they needed and would appreciate such opportunities. This is not to infer that drawing the Veteran into the life of the campus is an easy task. The group especially wanted to meet students who were not Veterans, both girls and men.

More than 60 per cent stated that they desired psychological testing, particularly with reference to vocations, although one out of five of them felt it might be very useful to aid him in a better use of his leisure time. One out of seven stated that he would like to be tested if the test results were interpreted to him. He indicated that students (and others) frequently had taken tests, but had seldom received adequate information about the results. Three out of four indicated that they would like to have a discussion or discussions with a counselor or faculty member, and one out of ten indicated that he had a personal problem with which he needed help. One-half of the group felt that they read too slowly, and two out of five felt that they needed aid with reading comprehension. Twenty per cent listed some seven different subject areas in which they felt that they needed aid. Mathematics and English were mentioned most frequently.

The avocational interests of these men were investigated and of the total group the most predominant avocational interest was in athletics, with one out of two indicating a desire to participate. The next two included cultural interests (two out of five), and handicrafts (one out of five). More than two-fifths expressed a need for individual recreation, while almost one out of three desired to have most or all of their recreation in group activities. More of those who had chosen Engineering expressed an interest in handicrafts. Those who had not made a vocational choice tended to be interested in passive avocational activities.

This group was asked to indicate the types of recreation



which gave them the most satisfaction. These results are indicated in the following tables:

TABLE 1

*Percentage Checking Interest in Avocational Activities  
(Possible to check more than one in each group. Numbers in tables are per cents)  
When being Entertained by Others*

Music .....	75.5	Theatre .....	46.0
Parties .....	62.5	Speech .....	21.5
Movies .....	58.0	Nat. Hist. ....	9.0
Dance .....	50.5	Art Exhib. ....	8.5
Others .....	6.5		

TABLE 2

*Entertaining Others*

Planning Social Occasions ..	41.5	Dramatics .....	10.5
Business .....	40.5	Vaudeville .....	10.0
Music .....	30.0	Drawing Cartoons .....	10.0
Directing Activities .....	15.0	Others .....	3.0

TABLE 3

*Outdoor Activities*

Swimming .....	70.5	Motoring .....	37.0
Baseball .....	56.5	Horseback Riding .....	36.5
Hunting .....	45.5	Boating .....	35.5
Fishing .....	42.5	Golf .....	32.0
Football .....	41.5	Hiking .....	30.0
Picnicking .....	41.5	Track .....	23.0
Tennis .....	40.0	Others .....	0.05

TABLE 4

*Miscellaneous*

Card Games .....	49.0	Pets .....	21.0
Photography .....	33.0	Pool .....	21.0
Radio .....	29.0	Chess Checkers .....	16.0
Group Games .....	23.0	Handicrafts .....	15.5
Others .....	9.0		

When being entertained by others, more than half of the men expressed a desire for Music, Movies, and Parties. From Table 2, When Entertaining Others, more than one-third were interested in the business arrangements and planning of social occasions. When reacting to out-door activities, more than half of them checked Swimming and Baseball, and in the miscellaneous

group, Photography was checked by one in three and Card Games by almost one in two.

Based on a study of 400 Veterans the following responses about their likes and dislikes in various school subjects were solicited. (They were not requested to check each course.)

The group stated that the sections of the newspaper to which they gave the most attention were in order: General News, Sports, Comics, Editorials. Fifteen portions of the paper were listed. Most of the men listed six to twelve sections which they read. Sixty-two per cent listed General News first, second,

*Subject Likes and Dislikes of 400 Washington University Veterans*

Subject	No.	% Like	% Dislike	Subject	No.	% Like	% Dislike
Algebra .....	292	54	19	Economics .....	144	33	3
History .....	240	49	11	Poli. Sci. ....	148	32	5
Physics .....	218	47	9	Phys. Ed. ....	184	31	15
Geometry .....	222	47	8	Music .....	184	31	15
Psychology .....	190	47	2	Distrib Forces ..	124	30	1
English .....	228	45	11	Biology .....	142	28	8
General Science ..	170	40	3	Composition ....	178	25	20
Chemistry .....	184	37	9	Sociology .....	108	24	3
Mech. Training ..	180	37	8	Philosophy .....	104	23	4
Trading .....	202	35	13	Accounting .....	116	22	6
Public Speaking ..	174	35	8				

third, or fourth; 41 per cent listed Sports in one of these four places; 44 per cent listed Comics in the first four places; 33 per cent listed Editorials first, second, third, or fourth.

In choice of periodical literature, more than four out of five stated that they read two or more magazines. Twenty-nine per cent of the whole group listed *Time Magazine* as their first choice; 22 per cent chose *Readers' Digest* first.

In choice of literature, 34½ per cent placed Fiction first, about 15 per cent preferred Science, and another 15 per cent, Biography, while about 10 per cent gave Adventure as their first choice.

When asked to check the occupations which they might be interested in pursuing, 58 per cent of the total group checked one or more of the occupations in the construction or mechanics area. Those with more than 10 per cent of the total group checking were: 1) Building Contractor, 20.5 per cent; 2) Radio,

18 per cent; 3) Aviator, 13 per cent; 4) Electrical Contractor, 11.5 per cent.

Fifty-two per cent checked the artistic area in this order: 1) Draftsman, 18.5 per cent; 2) Photographer, 14.5 per cent; 3) Architect, 13.5 per cent; 4) Interior Decorator, 12 per cent; 5) Musician or Music Teacher, 10.5 per cent.

In the scientific group, 57 per cent checked as follows: 1) Mechanical Engineer, 20.0 per cent; 2) Chemist, 14.5 per cent; 3) Mathematician, 14.0 per cent; 4) Aeronautical Engineer, 11.5 per cent.

Seventy-four per cent checked in the managerial and commercial areas in the following way: 1) Sales Manager, 18.5 per cent; 2) Advertiser, 15.5 per cent; 3) Certified Public Accountant, 14.5 per cent; 4) Retail Salesman, 10.5 per cent.

In the literary and humanistic area, 57 per cent checked: 1) Journalism, 19.5 per cent; 2) Radio Announcer, 15.5 per cent; 3) Physical Education, 15.0 per cent; 4) Teacher, 14.5 per cent; 5) Lawyer, 13.0 per cent.

The survey indicated a continuing strong bias toward Engineering and Business. This, coupled with an unrealistic attitude toward vocation on the part of many Veterans, indicates the need for better vocational information and more adequate counseling facilities. Interestingly, although many of these with a vocational choice showed a lack of realism, the group that showed itself to be most unrealistic were those who had as yet made no vocational choice.

It is felt that a regular study of the interests and attitudes of the students can make an important contribution to the teachers who have such students in class, to the counselor working with these individuals, and to those responsible for the welfare of the college as a whole.

# THE DIAGNOSIS OF READING DEFICIENCIES AS AN AID TO REMEDIAL WORK

FRANCES ORALIND TRIGGS

Chairman, Committee on Diagnostic Reading Tests

IN a society whose very structure is dependent upon its people for leadership, there should be concern for those who are handicapped to the extent that they cannot use reading as an efficient tool in learning. Even though reading is not the only avenue of formal learning, it has been in the past and probably for some time to come will continue to be, the most frequently used avenue.

There are two premises, I believe, on which we can start our consideration of diagnosis and its effects upon reading instruction: first, that there exists in our schools enough inefficiency in reading to be a cause of real concern on the part of the public and a cause for action on the part of the educators, and second, that a constructive approach to this problem can be made. Over a period of years research has shown the way; our problem is to absorb the lag between research and practice.

On the basis of a study of research in reading, a third premise might be stated: that all students do not profit equally well by the same approach to reading instruction. Or to put it another way, students profit when instructional techniques in reading are varied to their needs.

Interesting and significant in this connection is the fact that a large Child Guidance Clinic with which I have had some contact recently, finds that the students who go through their clinic and who have reading disabilities (but perhaps were not referred because of them) are in the upper and the lower brackets of general intelligence as measured by the Stanford-Binet Test. Few of them are in the large average group. This suggests that the reading instruction in public schools, generally speaking,

meets the needs of the "average" child, but that those who "deviate" at both ends of the ability scale, and who ought to have more varied instruction, do not get the help which they need.

On the basis of these three premises, first, that there is cause for concern because so many students do not learn to read efficiently; second, that techniques are known by which this condition can be improved; and third, that instruction varied to the need of the individual brings improvement in reading, it would seem that the next step in improving reading instruction is to make easily available to those who teach, tools which will point the way toward the types of instruction necessary to cultivate the essential reading skills. I do not want to make this sound too simple. Those who have worked with the students who are retarded in reading know that they cannot all be classified as simple cases of reading retardation. Some of them have emotional problems which must be removed or alleviated before anything can be done to help to improve their reading skills. In some cases, these emotional problems may be caused largely by reading disabilities; in other cases, an emotional problem, unrelated to reading, may be the major cause of the inability to read. It is well known that an emotionally unstable child does not learn anything so readily as the emotionally stable one.

A logical approach that a teacher might make in order to vary reading instruction in terms of student need, would be first to survey the group, to determine the level of reading skills of each individual and to compare that level to his ability and to the level of general development. For the students found to have reading skills average and above the level expected, the level of reading difficulty for class assignments could be adapted to the skills of these students using the data from the survey test as a guide. But for those students whose skills, on the basis of the survey test compared to their general level of development, seem to be totally inadequate to handle the level of material which is usually used for instruction, a different approach must be found.

It should be possible for this same instructor to use diagnostic reading tests with the students whose reading develop-

ment is found to be unsatisfactory to determine what skills are inadequate; in other words, why their reading development is retarded. From the results of such diagnostic tests, and other information concerning the individual, the instructor should be able to approach remedial instruction. With all students except for about 5 to 10 per cent of the retarded students, the classroom teacher could then do constructive remedial work. Those who did not respond to such instruction—the 5 to 10 per cent whose remedial instruction required a clinical approach, either because of emotional involvement or because of other special difficulties which might exist—perhaps would need to be referred to a reading clinic, if such is available, where remedial work would be carried on by intensive clinical methods to the place where these students could profit from regular classroom instruction.

At present, however, there are certain difficulties lying in the way of this "logical approach." A teacher can give a screening test to aid in determining the level of reading efficiency of individual students. She can also isolate students whose skills are seriously retarded. She can use well-known approaches in varying the instruction to meet the needs of those students whose reading skills are above average. There are a number of screening tests and other devices which might be used for this purpose. But the instructor gets few clues from the tests now available as to the appropriate form of remedial instruction for the more retarded readers. For instance, if on the basis of a survey or screening test, a student is found to have vocabulary skills below the level expected, poor comprehension of the usual type of textbook material, and the ability to read so slowly that he cannot possibly handle an assignment of the usual length in the time allotted, what can the instructor do for that individual and others like him in her class?

Persons experienced in remedial reading know that appropriate procedures cannot be planned unless or until the instructor has tests to further identify the reading deficiencies of such students. If diagnostic tools which also identified approaches to the remedial instruction were available, such instruction could be approached without having to depend exclusively on

clinical and individual techniques, and it would be expected that the number of students who reach the higher grades and colleges with reading deficiencies would be markedly reduced. Reading would then become a more dependable tool in a democratic society.

On the basis of this assumption, a group of specialists in reading at the junior- and senior-high-school and college levels reviewed remedial techniques which had been used successfully by persons working in the field. They found that in the following areas remedial techniques might reasonably be expected to be successful in improving reading skills: (1) attention to vocabulary, specialized to the area in which a student is reading, such as mathematics, physical, natural, and social science, and the humanities; (2) comprehension—obtaining the main idea, the supporting details, sequence of ideas, and conclusions to be drawn from a passage; (3) rate of reading—flexibility according to the content and purpose for which reading is being done, and the background of the individual doing the reading; (4) word recognition skills—ability to divide words into syllables and to sound them; to recognize prefixes, suffixes, and roots; and to hear and utilize sound and context clues in understanding the meaning of new words.

The committee of persons interested in this problem, therefore, met in May, 1945, after obtaining limited funds for their work, took up the task of developing tests in each of the areas named to help instructors to diagnose the deficiencies of poor readers, and from such diagnosis to approach remedial work.

Before describing the work of the committee in further detail, allow me to give the names of the members of the committee, to describe briefly the policies which they established, and to give recognition to the Blue Hill Foundation, the source of the funds which have made the work possible.

The names of the members of the committee are: Doctors Robert M. Bear, Dartmouth College; Ivan A. Booker, National Education Association; Daniel D. Feder, University of Denver; Constance M. McCullough, Western Reserve University; E. A. Monroe, Pennsylvania State College; George D. Spache, Chapqua Public Schools; Arthur E. Traxler, Educational Records

Bureau; and Frances Oralind Triggs, Educational Records Bureau, Chairman of the Committee.

The work of this Committee has been financed by the Blue Hill Foundation of New York, of which Mr. Fred Camp is President. Dr. R. O. Runnels, the Educational Consultant to the Blue Hill Foundation, has been the person who interested the Foundation in the work of the Committee on Diagnostic Reading Tests. The Committee would like to take this opportunity to publicly acknowledge the financial help which it has received from the Blue Hill Foundation, with special appreciation to Dr. Runnels, who has at all times helped the Committee in the interpretation of the project to the Foundation.

At its very first meeting the Committee determined that all of its work would be done on a non-profit basis, and that any money which might accrue from the sale of any materials which were developed by the Committee would be returned to the research, either on the tests themselves or on related projects in the field of reading. It was evident from the beginning that it would be some years before the basic research on the tests themselves was completed, and, of course, such work is costly.

The remainder of this report will describe in some detail the test which the Committee is setting up. I hope then that having described its work, there will be time for discussion of the project. I also hope that if there are those in the audience who would be interested in helping us to give these tests for norms, that they will see me after the discussion. There are probably several people in the audience who are already cooperating with us to this end,<sup>1</sup> but we are in need of cooperation of many such persons. I hope also that if there are suggestions for the Committee, you will take time either to write to me or to discuss them with me after this session. The Committee is very anxious to have such suggestions.

May I now describe in detail the tests which the Committee has and is developing and standardizing? The areas in which these tests fall have already been named: first, Vocabulary;

---

<sup>1</sup> Special mention should be made of the cooperation received from Miss Esther Ford, A. B. Hart Junior High School, and Mr. George Whitman, William Dean Howells Junior High School, Cleveland, who have obtained for the Committee all data on which original research on the tests was based.



second, Comprehension; third, Rate of Reading; and fourth, Word Attack or Word Recognition. The reasons for developing tests in these areas have also been given—they are the areas in which successful remedial techniques have been developed which an instructor can use in the classroom or in small group instruction. The *Manual* being developed for the tests will refer in detail to appropriate group remedial techniques which can be applied in classroom instruction for students whose reading achievement is found to fall in certain areas.

It is expected that the *Manual* will draw attention to pertinent portions of existing published materials which describe remedial techniques. Thus, a teacher using the tests and manual as a guide to remedial work, can feel confident that the initial procedures employed will be those which have been used successfully in other programs. With these procedures as a starting point, suggestions for changes and improvement will occur to the alert instructor, and thus modified procedures and new procedures will be developed.

It is hoped that case study material also can be used in the *Manual* for illustration.

The relationship of the interests of a student to successful remedial work is so important that this topic will also be treated in the *Manual*. Members of the Committee are considering the wisdom of developing an inventory which might be used for this purpose, though a formal inventory need not necessarily be used to guide the instructors in studying the reading interests of children. There are other techniques which have been found to be successful also.

The test battery on which the Committee is working is divided into two general parts, the *Survey Test* and the *Diagnostic Tests*. There are two forms of each test.

*The Survey Test:* The purpose of the *Survey Test* is to provide a reliable instrument for determining the general level of reading achievement for students in junior- and senior-high-school and college freshmen years. It provides scores on the general rate of reading with a comprehension check, a general vocabulary score, and a comprehension score on textbook type material. A complete item analysis has been made on this test; some portions, in fact, have been administered in two experi-

mental forms. The item validities and difficulties of each item have been determined. Judging from the administration of both forms of the test, and from an average of item validities and difficulties, the two forms of the test are quite comparable. The tests also have satisfactory reliability. Norms are now being established on this section of the test and it is expected that it will be ready for sale by June.

*The Diagnostic Tests:* The various parts of the *Diagnostic Test* battery are at different stages of development. Most of them are in a second or revised experimental edition and will go out for norms this spring. It may be possible to get some of them ready for distribution before September 1, 1947.

Section I of the *Diagnostic Test* battery is the *Vocabulary Test*. This test yields reliable scores in the following five areas: General Vocabulary, Vocabulary of English Grammar and Literature, Vocabulary of Mathematics, Vocabulary of Science, and Vocabulary of Social Studies.

Section II, Comprehension, is presented in two parts, Silent and Auditory. The tests measure comprehension on textbook type material drawn from the social studies, the sciences, and literary fields. The reading material is graded in difficulty according to the Lorge formula and according to other more subjective methods such as composition structure and type of content. The two parts of the test, silent and auditory, are comparable in order that it will be possible to compare the comprehension ability of an individual student when he himself reads the material silently and when it is read to him. Thus some estimate can be made of the extent to which a reading deficiency may be handicapping a student's ability to comprehend what he reads, or the extent to which insufficient maturity and poor background make it inadvisable to attempt to present ordinary textbook materials.

Section III of the tests, measuring Rates of Reading, has three parts. The first part of the test measures the student's usual rate of reading with comprehension interesting, story-type material, as compared to his ability to adapt his rate of reading, and still comprehend comparable material, when he is instructed to vary his rate. On both parts of this test a check is made on the student's comprehension of the material read.

Part 2 of Section III measures the rate at which the student usually reads and comprehends social studies material.

Part 3 of Section III measures the student's usual rate of reading, with comprehension, science material.

These areas correspond to three of the areas of vocabulary measurement in Section I.

Section IV of the test is entitled Word Attack. It measures word recognition skills and is divided into two parts, Oral and Silent. The Oral Test is the only individually administered portion of the battery. Six graded paragraphs of interesting, general-type reading material are read aloud by the student. The teacher marks errors indicating in each case whether it is a repetition, substitution, omission, or mispronunciation, using a modified method of scoring based upon that of the Gray Oral Reading Paragraphs. The establishment of grade norms as well as percentile norms and research on the relationship of types of errors in Word Attack as measured by Part 2, the silent section of this test, and to scores on other parts of the test, represent interesting research possibilities. The test-retest reliability on total errors on this test for a group of approximately 100 ninth-grade Cleveland students was .94, when repetitions were weighted one to each phrase, and when weighted one to each word was .90. Spearman-Brown reliabilities were even higher. Research on remedial work following diagnosis in this area has hardly been touched.

Section IV, Part 2, Word Attack, Silent, attempts a new approach to the measurement of word recognition. This part of the test is set up for group administration. Through it we have attempted to measure the ability to hear sound by using a matching technique, and to test the student's ability to divide words into syllables. We know that remedial techniques along these lines can be used profitably to improve the reading skills of some students and that remedial exercises are available which classroom teachers can use. We are not sure how such skills are related to the total reading process but this is being studied. We do know the instrument which we have built is reliable.

A number of experimental studies are under way on the use of these tests in varying situations. For instance, to test their ceiling, they are being given to selected groups of college stu-

dents. To determine whether it is necessary to read the material to obtain a good comprehension score, comprehension questions are being asked without giving the subject an opportunity to read the material on which the questions are based. Norms are being set up on well-defined groups in wide geographical areas, in all of the grades from the seventh through the college freshman year.

The method by which these tests will be distributed has been carefully considered each time the Committee has met. The project represents a departure from the usual procedure followed in test construction. The only test in the series which is definitely adapted for use within a single class period is the Survey Test. The Committee has chosen, in the beginning at least, to strive to develop the most valid and reliable instruments possible rather than to make the tests conform to any stated time limit, such as a class period, or other arbitrary administrative pattern. The diagnostic battery usually will be used only for limited numbers of students in any one school. Therefore, it should be possible to adapt it to the time requirements. As research progresses on the test, the Committee wants to be free to make changes and to adapt the tests according to research findings.

All funds accruing from the sale of the tests will be put back into the fund for further research and for such revisions as seem warranted. Our real problem now is to finance the first printing and distribution of the tests.

The Survey Test will be ready for final printing by June 1. Tentative norms will have been established and will be available for interpreting scores. The first edition of the test probably will be copyrighted and printed by the Committee itself and they will very likely be distributed by some non-profit educational organization. Modifications in the tests will then be made as research on them and experience with them as guides to remedial work accumulates. Thus the Committee hopes to contribute techniques which will help to absorb the lag between research and practice in the development of more efficient reading instruction and to reduce the number of students who are identified late in their formal education as needing remedial reading.

## AN IN-SERVICE TRAINING PROGRAM FOR RESIDENCE COUNSELORS

DUANE COLLINS.

Assistant Professor of Education, University of Connecticut

DR. COLLINS described the in-service training technique which he found useful with residence counselors at the University of Connecticut. At the University of Connecticut there has been a phenomenal increase in the number of students from 1000 to 7000. This has resulted in the hiring of a large variety of counselors, few of whom ever had any real training in counseling procedures. As a stimulation to professional growth and competence, and as a means of increasing the motivation of the counselors to improve their work, a book-passing technique was adopted. Books concerning counseling practices and procedures were purchased individually and distributed to the counselors. Dates for passing the books on to still other counselors, and dates for discussion periods were set.

The total list of books was divided into three levels: (1) those which the counselors were required to read, (2) those which it was recommended as desirable to read, and (3) a list from which the counselor could select any book he wished.

The results of this procedure, Dr. Collins believes, were as follows: (1) the professional growth of the counselor, including an increase in the knowledge of terms, concepts, philosophy, and diagnosis of personnel work; (2) an improved morale in unity of purpose among the counselors; (3) a setting of the stage for the clinical case conference approach to student problems.

# THE USE OF DORMITORIES FOR SOCIAL EDUCATION

FLORENCE THOMPSON

Dean of Women, Michigan State College

As an introduction to the problems of social education I should like to refer you briefly to an article in *Fortune Magazine* for November, 1946.<sup>1</sup> It is an article about Elton Mayo, of the School of Business of Harvard College, whose research in industry is widely acclaimed. The central theme of the article is a proposal to management and labor of a social basis for industrial peace. It was shown that Mr. Mayo's devotion to scientific truth in industrial experiments led him to alter his entire axis of thought from the proposition that all social problems are essentially individual to the proposition that all individual problems are essentially social. It was pointed out that man is a social animal and that the material aspects of our civilization have changed through industrialism while man unfortunately has not. Like John Dewey and others, Mayo notes that step by step with our economic progress there has been a destruction of individual significance in living for the majority of our citizens. The issue the civilized world is facing now is a rapid industrial, mechanical and physico-chemical advance so rapid that it has been destructive of all the historic personal and social relationships. While material efficiency has been increasing, the human capacity for working together has in the same period continually diminished. Our methods are pointed at efficiency, not at the maintenance of cooperation. We know how to devise efficient methods; we do not know how to ensure spontaneity of cooperation. Mr. Mayo's research indicates the necessity for understanding and cooperation among individuals

---

<sup>1</sup> Anon. "The Fruitful Errors of Elton Mayo." *Fortune*, XXXIV (1946), 181-183 +.

if we are to have industrial peace. That generalization also would hold in the matter of peace among nations and peoples.

If it is true that individual problems are essentially social, is not training in social education a suitable objective for higher education? What facilities do higher institutions have for such training? Are the higher institutions making the best possible use of the facilities they have for this important branch of training?

It was my pleasure recently to study several aspects of residence-hall programs in this country. The primary purpose of that study was to find out what possible contributions the residence hall might make to the entire educational program. Results of that investigation<sup>2</sup> indicate that the hall can be expected to make its most valuable contribution in the area of human relationships.

Data in regard to this aspect were gathered from interviews with groups of students and with head residents during personal visits to several representative institutions. For purposes of analysis the material was classified into the following four areas of relationship outlined by the Progressive Education Association<sup>3</sup> and based on the needs of youth in American culture. They are: Personal Living, Personal-Social Relationships, Social-Civic Relationships and Economic Relationships.

In the area of Personal Living some of the learnings students believed they experienced were: (1) the building of a satisfactory personal philosophy, (2) the formation of individuality through self expression and in turn the integration of personality, (3) emotional maturation including emancipation from the family, (4) the acquisition of knowledge, (5) the development of interests, appreciations and attitudes, and (6) the development of skills and techniques.

The area of Personal-Social Relationships includes the immediate face-to-face situations, and learnings in this area were said to include: (1) social customs and techniques, (2) the

<sup>2</sup> Thompson, Florence M., "Provision for Student Activity Programs in College Residence Halls for Women." Unpublished Ph.D. dissertation, Department of Education, University of Chicago, 1946

<sup>3</sup> Progressive Education Association, *The Social Studies in General Education*, pp. 116-237. A report of the Committee on the Function of the Social Studies in General Education. New York: D. Appleton Century Company, 1940

formation and evaluation of friendships, and (3) the ability to get along with all kinds of people.

Social-Civic Relationships are considered to be those out beyond the face-to-face contact. In this area they learned: (1) techniques for working with groups, (2) how to be leaders and how to select leaders, (3) the meaning of being responsible to the group or for the group, (4) consideration for individuals and groups, (5) the necessity for appropriate regulations for group living, (6) the importance of cooperation to a smooth-running enterprise, (7) loyalty to the group, (8) tolerance, and (9) freedom.

The area of Economic Relationships was seldom mentioned. This fact might be accounted for largely because the study was made when money was plentiful. In only one residence hall studied intensively was any definite plan in effect for giving students practice in their chosen vocational fields, such as dramatics, personnel administration, library, journalism and home management.

The four areas as outlined are not mutually exclusive. Needs met in any one area are influential to more satisfactory relationships in the other areas, resulting in a reciprocal effect.

In order to give you a "feel" for the benefits which students believe it is possible to derive from their residence-hall experiences may I quote a few verbatim statements.

It's an inspiration for further achievement to be with others who are interested in further learning.

If one girl is a musician it "rubs off" on others.

You see new interests but they are "contagious" only if they fit your personality.

You become aware of so many things and it gradually grows; you're aware of little things.

You learn good taste in dress. Some one says about your favorite dress, "Huh uh."

Here in the residence hall you can start with a clean slate and make anything you wish of yourself—with your home friends, "No."

You change your standards of success. With increased competition you learn not to try to do and be the best, but to do your own best.

You learn emotional control—you learn that people will stand for just only so much mood and temper.



In hashing over problems you learn that you aren't the only one who has them. This minimizes your own.

Other parties are more worthwhile than a dance because they call for the use of experiences and techniques to which you aren't as well accustomed.

You learn how to serve tea, greet chaperones, and talk to faculty members—as a freshman you're petrified, but as a junior you're at ease.

The working relationship is important—you have to be doing something that matters in order to work up loyal friendships. You discriminate between the people you like and the kind that's better for you; you can't run around with two hundred and fifty.

You become less selfish about the other person's good time. You are glad to help a girl dress for a "formal" even if you aren't going. Maybe she's even going to wear your dress.

We do everything here by student opinion, so every person here must be trained to be a potential good citizen.

When you're made responsible for a certain job you feel more like you "belong" from then on.

Any girl considers her opinion as important because she's always consulted.

You learn that each individual is important. The opinion of every person in the hall must be considered before the group is committed to any special action.

We learn how to select leaders and they aren't always our best friends.

You learn to get things done *on time*.

A dorm doesn't run by laws laid down but by individuals taking personal responsibility.

In regard to regulations, you *do* something about a situation, you don't just talk.

Maybe you're a Baptist and you go to church with a Methodist and you see that it's a pretty good church.

Just as the *end product* of learning experiences was found to be largely in the areas of human relationships, so was it also evident that the *process* of learning was rooted in dynamic social forces. That process takes place in two general ways. First, through *social interaction*—students learn simply from contact with many other minds. Second, they learn definite procedures and techniques through *working* with others in activities. It might be assumed that what is learned through the first method would not be at all dependent on the second.

The investigator was careful to inquire about this point and found it to be the unanimous opinion of the students that learning through social interaction was more productive and effective when supplemented by the activity program. The work in these activities vitalized their personal contacts and therefore supplemented the first method. Students believe that the intimacy of their contacts is a strong determinant in the first method and that the working relationship is likewise effective in the learning of procedures and techniques.

The quantity and quality of this end product, social education, varies considerably from hall to hall. An attempt was made to find out what is responsible for such variation—what some of the factors are which influence the nature and caliber of the program. It was found that one certain relationship stood out above all others: wherever there was a “good” program, a high degree of *esprit de corps* was also evidenced. Close analysis would indicate that this finding should have been anticipated. When it is remembered that “social interaction” and “a working relationship” were basic to the learning process, it might be expected that the attitudes and sentiments of the group, individually and collectively, would be highly influential in determining the end result.

Interestingly enough a parallel to this condition was also found in industry by the Mayo group in their studies of the productivity of workers in the Western Electric plant. All experiments dramatically and conclusively demonstrated the importance of employee attitudes and sentiments. It was clear that the responses of workers to what was happening about them were dependent upon the significance these events had for them. This discovery is nothing very new or startling. It is something which anyone who has had some actual experience in handling other people intuitively recognizes and practices.

This spirit which is conducive to the formulation of a program based on friendly cooperation, is very evident in some halls. In those halls, students were profiting immeasurably from the program and were well aware of its benefits.

From the facts presented it is evident that dormitories make possible important educational services. It is also true that in

some institutions their purpose is considered to be primarily one of shelter. It seems surprising that some administrators are willing to spend thousands of dollars on the erection of beautiful buildings and in turn give such little thought to what goes on inside them.

Reflection upon the relationship of *esprit de corps* to effective social education emphasizes the importance of the adult leader in the situation. Any attempt to delineate those traits which combine to make a "good" head resident would be as difficult as to describe the traits of a "good" teacher. The head resident must basically be able to work with people. She must also have a conception or at least an intuitive feeling for the objectives of residence-hall living. With such natural endowments she can accomplish much. She can accomplish more if those natural endowments are trained through formal education. Such training should increase her ability consciously to recognize appropriate educational objectives and to plan and execute a program accordingly. Some of the aspects she will need to consider along with members of the group are:

1. Experiences which are worthy of inclusion in the program.
2. Suitable bases of evaluation.
3. Areas in which students are best fitted to assume responsibility and a proper progression in the amount they assume.
4. Extent of formal rules and regulations.
5. Provision of opportunities for students to participate in a manner and to an extent which is in keeping with their needs.
6. Appropriate techniques for encouraging participation.

A second factor observed to be influential in the establishment of *esprit de corps* was the size and lay-out of the living unit. In very large halls it is extremely difficult to achieve social orientation and development because they normally originate in smaller groups and gradually work to larger ones. Individuals are apt more keenly to feel their responsibilities when the number is not so large as to create anonymity.

Administrators would do well to study living units in relation to their educational advantages before building student residences. They need also seriously to consider their obligation to select competent head residents. Such considerations point out a possible way of bringing the residence hall nearer to the center of the field of learning experiences rather than on the periphery where it now tends to be.

## COORDINATION OF STUDENT PERSONNEL SERVICES AT CORNELL UNIVERSITY

LUCILLE ALLEN

Counselor, Cornell University

DR. LUCILLE ALLEN described the coordination between various student personnel services which is occurring and which is planned at her institution—particularly that between her office and the Health Services. Dr. Allen indicated that next year it is planned to have five patrons in each residence—3 men and 2 women. These patrons will be faculty staff members, they will have their meals at the residence once a week and through their contacts with the girls will discover what the girls want and need. Dr. Allen felt that such a relationship between the patrons and the students would be mutually helpful and would result in the living units being used to the fullest extent. New living units, which are being built, are in the mood of this age, i.e., they are somewhat modern and truly functional.

The coordination of counseling services and health services was aided by Dr. Allen's being invited to become a member of five faculties at Cornell, including the Medical Health Service. This condition was extremely helpful in getting the students to actually use the services.

A joint appointment was made between Dr. Allen's counseling services and the Medical staff. Miss McCormick, who has her M.A. in Nursing Education and who, in addition, had five years' experience in a psychological clinic and with a psychiatrist, was chosen for this position. Miss McCormick goes into the dormitories for any reason and makes direct contacts with the girls. If an emergency arises she is empowered to act at once.

If a difficult problem arises, the girl is referred to Dr. Allen's office where a tentative diagnosis is made. Usually the psychia-

trist provides additional information and instructs Dr. Allen's office to carry on with the treatment.

Dr. Allen expressed the belief that the close relationship between her office and the Medical Service will have a real effect in preventing epidemics and in increasing the respect for good health.

Dr. Allen then described a student-dean program. There are ten assistants, each of whom spends one year as head of a cottage, and one year as an assistant in the dormitories. These assistants are assigned functions such as organizing music hours, etc., which the University can not provide formally. Such procedures broaden the opportunities of the whole student body.

## CURRENT TRENDS IN PERSONNEL WORK AT PRATT INSTITUTE

MARGARET STEPHENSON

Director of Student Life, Pratt Institute

WHEN Dr. Lloyd-Jones asked me to participate in this discussion tonight on "New Trends in Personnel Work" I assented but told her that I did not think there was anything *new* in what we are doing but rather that we are using tried methods in a new situation. However, there is one angle of the program which is unusual—that I, a woman, head the personnel program for *both* men and women (the ratio is 2 men to 1 woman). There is a man and a woman counselor on the staff besides myself. Those persons I happen to know who carry the title "Dean of Students" in co-educational institutions are *men*. I doubt if you can call having a woman in this position a new "trend"—but rather an atypical situation.

Can you imagine what it would be like to step into a brand new job where there are no patterns set for personnel work? No one whose place you take? It was a tremendous challenge to me. When I accepted the position I had visions of not having to bide my time to make gradual changes in the work of my predecessor. However, it did not take me long to find out that it is *just as difficult*—or even more so—to bring about changes from a *laissez-faire* policy as it had been in my two previous positions to make changes from policies set up by my predecessors. In fact, many times when I have felt frustrated in trying to work out standards I have realized how in previous positions which I have held—similar to those held by many of you—that someone long before me had fought and bled for those standards which I had blithely accepted.

The policy of the Institute concerning student welfare is summed up in the statement which appeared in the catalogue

as late as 1944-45. It read, "Prospective students should be sufficiently mature to be responsible for themselves." Because the philosophy of the Institute was changing, our office was set up September 1945.

Pratt Institute is a technical college made up of four schools—Art, Engineering, Home Economics and Library Science. The last named is not an undergraduate school and so does not belong in the same category in our plans.

One of the first steps in getting acquainted with the faculty and the school was to make a study of the backgrounds of the student body—including the age range, the number living at home and commuting, how many hours they spent daily in travel, and the number living in rented rooms near the college. As we got acquainted with the set-up we decided that it would be better to put our emphasis for the first year on group guidance rather than on individual guidance. This decision was due partly to the urgency of the need for group guidance as expressed by student and faculty, and partly because we could work in this area more safely without treading on the toes of the persons who had been doing individual guidance for years, and although untrained, had in many instances been doing a good job.

Therefore early last year we set up long-range planning committees—a Club Committee, a Social Committee, a Finance Committee and a Committee to Revise the Student Government Association. There were 5 students and 2 faculty members on each. These committees are still functioning—in addition to some new ones including a Publications Committee. It has been their job to make policies concerning the activities which, like Topsy, had "just grown" in a lackadaisical fashion. A few had flourished but most of them, because of lack of direction, had merely existed. Now through the work of these committees we have a system for chartering clubs, for registering and controlling social affairs, for budgeting and auditing the student activity fee, for policies in choosing the editorial staff of the newspaper and *Yearbook*, for policies and regulations not very different from those you are accustomed to on a campus where there has been student government for many years—the



kind of thing which I had taken for granted when I was at the University of Oklahoma without realizing that someone a long time before me had had to do a lot of ground work.

Because the three undergraduate schools at the Institute are run on the basis of strong States rights rather than by a Federal Government—the student body had its primary loyalty to its own school rather than to the Institute. Thus when the new Student Government was set up it was deliberately planned to cut through school lines horizontally with representation to an all-Institute freshman-class council, sophomore-class organization, etc., instead of an organization of mechanical-engineering freshmen or costume-design sophomores. Last spring we had our first all-student-body elections. It has meant careful planning every step of the way—but we are getting places

In the area of the social program we have made considerable headway. The students wanted a hangout—a place where they might drop in casually—to smoke, to dance, to chat. As a result of a Winter Carnival they cleared enough money for a start. This fund has been added to in one way or another. A committee made up jointly of students from the Interior Design class in the Art School and from the Home Furnishings class in Home Economics rounded up some old furniture and painted it. They put a juke box and a coke machine into what had once been a gymnasium. It was opened in December, 1945. The response was very gratifying. In the year since then about \$33,000 has been put into new furnishings and in remodeling a room off the lounge for a snack bar. A hostess is on duty from 11 a.m. until 9 p.m. We now average 1500 persons a day using the lounge. It is so crowded that they play bridge on the floor and eat hamburgers standing up. Instead of buying decks of cards and checkers and chess men by the half dozen, we now order them by the gross. We give out tickets for radio broadcasts, organize theater parties for Broadway productions and I could go on and on with examples. The point is that we have been able to meet a very definite need of the students and we already have in writing plans for a Student Union Building as soon as construction is possible.

Before I leave the subject of the social program, I want to

say that we are trying to meet the need for low-cost recreation on weekends. If a student wants to spend \$15-\$25 on a date, he has all of the resources of New York City. We do not need to compete with big name bands and night clubs. Our problem is to furnish a place for our 600 students who rent rooms within a few blocks of the Institute. So we are keeping the lounge and snack bar open on Friday and Saturday nights until midnight with all kinds of activity going on. They can have a wonderful time spending only nickels—or spending nothing as far as that is concerned. Besides the activities in the lounge, the Junior Class is sponsoring low-cost movies. Thus we are attempting a real campus-centered social program in the heart of New York City. We feel that we are furnishing some competition to some of the cheap places in the neighborhood which were student hangouts.

In the area of housing we feel most frustrated—as probably do most of you. We had no dormitories and women who formerly rented rooms to students prefer roomers who are in business and who can pay rent twelve months out of the year. We have far more students than normally and far fewer houses than before the war. In a city college there is no green grass and open space in which to spread out with temporary housing and New York City will not allow such units as Quonset huts. However, we have furnished a mansion—and I mean mansion—which belonged to one of the Pratt family. We have 33 freshman girls living on the second and third floors and we use the main floor for our Women's Club program. We were able to get government aid that included one dormitory unit for 80 veterans and 16 units for married veterans—thus using a large part of what small campus we had. It is a drop in the bucket, but it helps. A Faculty Housing Committee is working very hard, but with cramped quarters in a big city there are not the ways of expanding housing that there are on a residential campus.

We have made noticeable progress in the area of health. For many years the Institute had maintained an infirmary with 2 resident nurses—but they kept office hours at the Institute only an hour a day. As you well know, a person cannot choose

a particular hour to be ill or to have an accident—and the students were not making much use of the infirmary because it was  $4\frac{1}{2}$  blocks away. So we felt that the core of the health program should be centered in the main building rather than in the infirmary. A doctor who is a specialist in making surveys of health set-ups was called in to make recommendations. As a result we now have a Health Service in the main building with a doctor and 2 nurses who keep office hours from 8:30–4 p.m. and a real health program is underway. Needless to say the students are finding their way to the office in great numbers.

I could go on at length but I want to save time to tell you that we are not neglecting the area of individual guidance—which is the main interest of the three of us on the staff—but we have had to go slowly. Up to now we have given help to individuals who have found their way to us of their own accord—through knowing one of us while working on a committee, or by chance. However, they are coming to us in increasing numbers. What is more, we are beginning to get some referrals from faculty members and the deans—which is indeed a victory. It took a year before we could get the administration to appoint a committee to plan for an over-all guidance program. But we have it now and the objectives are to evaluate our present program, to see what areas need strengthening and how we may integrate and expand our personnel services.

I have hit only a few high spots but I have tried to give you an idea of what we have been able to accomplish in a year and a half with a new program. You see, these are not new trends or new methods—except as they are applied to this situation. I have been accused by faculty members who were satisfied with the old status quo of trying to fit a liberal arts and a residential campus pattern on a technical college with 60 per cent commuters. And I have had moments of questioning that too. But after this experience I believe even more strongly that we have the same obligation to provide personnel services for our kind of student body as you do on a residential campus—although we do it differently. We have exactly the same responsibility for the kind of men and women who graduate.

## THE NATIONAL STUDENT MOVEMENT AND CONFERENCE

JOHN BERGSTRESSER

Assistant Dean of Students, University of Chicago

At such a professional conference as this one, it is natural for us to place great emphasis upon the improvement of personnel techniques, measuring instruments, and administrative organization—and all of this is both fascinating and vital. This type of emphasis does, however, make it necessary for us to guard against the danger of thinking of students merely as rather passive customers *to whom* or *for whom* we are eternally doing something—such as, administering and interpreting tests, dispensing information and wisdom, arranging activities and housing, and so on. During these professional conferences, perhaps some (and I include myself in this number) tend to forget or to underestimate what Don Shank in his paper called “the dynamic interests of a modern student body.” Or, to put it another way, at times we may forget or underestimate what Ed Williamson in his paper spoke of as “the societal aspects or implications of student personnel services and techniques.”

In the light of this introductory comment, I hope in a very few minutes to arouse your interest and curiosity about an embryonic national movement which is almost completely student initiated and directed, and which is potentially of very great significance to higher education in America, especially to those of us who are engaged in college personnel work.

Underlying this student movement, it seems to me, is a nation-wide ferment in student thinking and a mounting concern about the student's own role in the shaping of his own present and future—both as college student and as citizen in the atomic age. One straw in the wind which has pointed to this trend in my experience has been the receipt of dozens of

letters coming from student leaders and professional colleagues in every section of the country, including Alaska and Hawaii. I refer to letters which have asked for information and suggestions regarding the establishment or the strengthening of student government.

But this concern about the student's role in the drama of modern society—and especially in that part of it which is played on the college campus—is apparently developing into more than a local or even regional phenomenon.

Last summer a group of 25 American students representing several colleges and several national student organizations went to Prague to attend an international student conference. One of the by-products (probably the most important one) of this venture was a decision by these American delegates to attempt to bring into being on their return a large, representative, non-political, truly national student organization in this country. They had in mind an organization somewhat similar to those already existing in such countries as Sweden and England, but an organization which is thoroughly adapted to American conditions.

As a first step, the group of 25 students set out last fall to organize a national student conference to consider (1) whether there is sufficient interest and desire among students nationally to undertake such a project and (2) if so, upon what general principles should such a national organization be founded.

Definite invitations to the colleges and to many of the national student associations of the country to send delegates or observers to a conference to be held at the University of Chicago on December 28–30 were not issued until November. Preparations for the conference were severely handicapped by financial limitations and lack of adequate publicity, work space, and clerical help. Those who planned the conference hoped, in spite of all these limitations, that 250 to 300 students would attend. When the conference registrations were tallied, however, there were 500 official delegates and 227 accredited observers, making a total of 727 registered student participants.

The delegates represented 297 colleges and 16 national associations. The observers came from 10 other colleges and 12

additional national associations. Thus 307 colleges and 26 national associations were represented. This was a response far beyond the most optimistic hopes or estimates—a truly remarkable demonstration of the widespread interest among today's students in the proposals to be discussed at the conference.

Because this conference was so completely student initiated and directed, I wish that we could have heard directly from one or more of the student leaders about the results of the conference. In fact, I tried hard to sell this idea to the Program Committee of the C.G.P.A. and A.C.P.A. (Personally I believe that in general we are overlooking a good bet in not inviting a moderate degree of student participation in our programs.) Since we cannot hear directly from the student leadership, however, I welcome the opportunity to give you the following abbreviated and general statement about the outcomes of the conference. These facts are taken from the printed report prepared by the National Continuations Committee set up by the Chicago Conference. (Copies of this report have been mailed to all members of A.C.P.A. through the courtesy of the Continuations Committee and the good services of our secretary, Thelma Mills.)

The Constitutional Assembly is being planned for early September in Madison, Wisconsin. The Continuations Committee is setting its sights on an attendance of 2000–2500 from nearly 1000 colleges. If they come even close to these figures—and proceed to create a national student organization along the lines agreed upon at Chicago—then, I submit, this student movement will have developed into something too big and potent to be ignored by college personnel workers. My plea is that we take the trouble to become thoroughly informed about this movement, try to understand its causes, and attempt to predict and evaluate its implications for student personnel purposes and programs.

# THE KEUKA COLLEGE FIELD WORK PLAN

KATHIERINE BLYLEY

President, Keuka College

PRESIDENT Blyley described the study and work organization which is in effect at Keuka College. This program grew out of research which began 8 years ago for some technique which would relate college studies to the real life situation. In June of 1940 there was a faculty workshop session at which the whole college year was rearranged so as to provide for 7 work periods in the field. During the freshman year this field work consists of a critical commentary on an assigned reading program. During the sophomore year the student spends part of his time in some community service project. During the junior and senior years the students spend a part of the year securing actual vocational experience.

Each student has a work counselor, a member of the employing agency, who trains and supervises the student and who makes a rating which is returned to the college.

When the student finally returns from his field work there is an evaluation day during which students meet with the appropriate departments, make oral reports and then have a conference with their major adviser. A grade is finally prepared and reported to the Registrar. This procedure provides then for both a study and a work grade record.

The community and service aspect of the program during either the summer or winter period is possibly the most unique aspect of the program. The student takes part in some worthy community project. He is supervised on this community job and is visited by the college representative.

Keuka College believes that education is better served if the student spends a specific period of time during each year away from the campus in supervised work or other types of

experience. The College accepts the responsibility of seeing that every student utilizes her time in a purposeful program designed to fit her as a responsible worker and citizen.

I. Objectives:

A. Over-all

1. To link education more closely with life experience, thus avoiding the artificiality of an education exclusively based on theory.
2. To provide through experience practical as well as theoretical vocational guidance.
3. To activate the maturing process by utilizing these experiences for the growth and enrichment of the total personality.

B. Objectives of Reading Commentary (Freshmen)

1. To develop English for use.
2. To train students in the use of the elementary tools of research.

C. Objectives of Community Service (Sophomore)

1. To develop insight into the community by participation in one community activity.
2. To develop appreciation of responsibilities as citizens.

D. Objectives of Vocational Experience (Junior and Senior)

1. To Provide practical work experience.
2. To provide first-hand observation, investigation, and exploration.
3. To give a tryout opportunity in the field of professional choice.
4. To become the focal point of the vocational guidance program for each individual student.

E. Objectives of Elective Experiences:

1. Earning
  - a) To aid in providing for student's college expenses.



- b) To give the students the experience of securing and holding a job.
  - c) To give the student an introduction to the world of work.
2. Creative Arts
- a) To enrich and widen the student's cultural experience in situations other than those which the college can provide.
  - b) To provide opportunity for experimentation for those with special artistic talent
3. Research
- a) To stimulate independent study in a chosen field.
  - b) To give the undergraduate foundation for further study.


#### F. The Relation of the Field Period to the Academic Program:

The relation of the Field Period to the academic program of the college has been well summed up by Acting President Blyley in the original announcement concerning the Field Period Plan, in which she indicated that whatever form of activity was chosen by the student, the aim would be "to bring the student into close and uninterrupted association with some aspect of life or learning which will serve to sharpen and direct her perception of values and give added meaning and significance to her regular academic studies."

#### II. A. Requirements:

- 1. Reading  
Commentary 1 credit 3 weeks
- 2. Community  
Service 1 credit 3 weeks
- 3. Junior  
Vocational 2 credits 5 weeks (uninterrupted)
- 4. Senior  
Vocational 2 credits 5 weeks "

## B. Electives:

- |                  |           |                         |   |
|------------------|-----------|-------------------------|---|
| 1. Earning       | 1 credit  | 3 weeks                 |  |
| 2. Creative Arts | 2 credits | 5 weeks (uninterrupted) |   |
| 3. Research      | 2 credits | 5 weeks                 | "   |

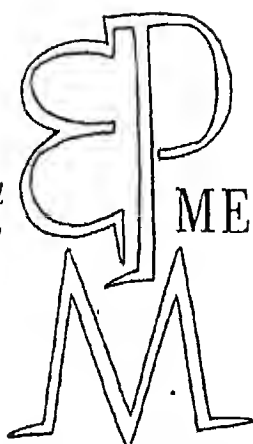
## III. Time: Winter Field Period:

November–December

Summer Field Period:

June–August

EDUCATIONAL and  
PSYCHOLOGICAL



MEASUREMENT

**VOLUME SEVEN, NUMBER FOUR, WINTER**

<i>The Effect of the Nature of the Criterion upon the Validity of Aptitude Tests.</i> DEWEY B. STUIT .....	671
<i>Relationship between Military Occupational Specialty and Army General Classification Test Standard Score.</i> NAOMI STEWART .....	677
<i>The Validity of a Multiple Aptitude Test at the Secondary School Level.</i> DAVID SEGEL .....	695
<i>Student Personnel Work in National Setting.</i> C. GILBERT WRENN .....	707
<i>Measuring Outcomes of Classroom Therapy.</i> HAROLD B. PEPINSKY .....	713
<i>The Frequency of Errors in the Classification of Individuals on the Basis of Fallible Test Scores.</i> ROBERT L. EBEL ....	725
<i>The Validity of Standard and Custom-Built Personality Inventories in a Pilot Selection Program.</i> DONALD E. SUPER .	735
<i>The Graduate Record Examinations.</i> K. W. VAUGHN .....	745
<i>Validation of the Study of Values for Two Vocational Groups at the College Level.</i> HAROLD G. SEASHORE .....	757
<i>Entrance Examinations at the City College of New York.</i> LOUIS LONG and JAMES D. PERRY .....	765
<i>Selecting Executives by Psychological Tests.</i> CLAUDE EDWARD THOMPSON .....	773
<i>A Misuse of Group Tests of Intelligence in the School.</i> D. A. WORCESTER .....	779
<i>Multiple Forms of Objective Tests: A Test-Scoring Machine Technique.</i> VERNON W. LEMMON and PHILIP H. DUBOIS .	783
<i>Book Review.</i> CHARLES I. MOSIER .....	787
<i>Measurement Abstracts</i> .....	791
<i>The Contributors</i> .....	801

## COMPLETE BACK VOLUMES NOW AVAILABLE!

In response to a considerable demand for out-of-print copies of EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, Volumes I through IV have been reprinted in a small-print edition which is available at \$3.00 per volume (paper-bound). The regular page size has been retained, so that these volumes match the regular volumes except in thickness. Four pages of the original have been reduced to one page of the small-print edition. The reduced print is readable, and will be found to be quite adequate for most research purposes. A sample page will be sent on request.

Volumes V, VI and VII are available in the regular edition at \$5.00.

Please send orders to EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 917 Fifteenth Street, Washington 5, D. C.

# THE EFFECT OF THE NATURE OF THE CRITERION UPON THE VALIDITY OF APTITUDE TESTS<sup>1</sup>

DEWEY B. STUIT  
State University of Iowa

It is generally agreed by personnel research psychologists as well as by guidance workers and counselors that validity is one of the most important, if not *the* most important attribute of a test. In selection, classification and guidance we are interested in predicting the future performance of an individual on the basis of what we know about him at a particular moment. This means that we must ascertain the relationship between certain measurable attributes and subsequent performance. Knowing the nature and magnitude of this relationship we can with more or less success predict the individual's later performance from his performance in the predictor variable.

Unfortunately the determination of this relationship is far from a simple proposition. There are at least three general factors which influence the magnitude of a so-called validity coefficient. First is the matter of the test itself; for example, a test of low reliability has automatically placed upon it a ceiling above which its validity cannot rise. As several recent writers have pointed out, reliability is no virtue in itself but there is no denying the fact that a test of very low reliability cannot have a high validity. In general, therefore, it is in the interest of validity to have a test of reasonably high reliability.

In addition to a test's reliability there are other important statistical qualities which determine the goodness of the test, e.g., the intercorrelations of the items. Time and space are not available in this paper for an elaboration of each of these points

---

<sup>1</sup> A paper given at the meetings of the American College Personnel Association, Columbus, Ohio, March 29, 1947.

except to say that there are many factors which influence the quality of a test—and all in turn may have a significant effect upon the test's validity.

A second factor which influences the size of a validity coefficient is the nature of the experimental population. All too often test manuals simply report that the validity of a particular test is a certain magnitude without describing the population from which it was derived. Was it an unselected population or was it one which had been selected on some basis, thus reducing its range of talent? Was the population representative of the universe or was it biased in some manner? What about such factors as education and experience? Was the population homogeneous with respect to these factors or were these factors shown to be unrelated to the traits under consideration? The careful user of tests will be diligent to inquire into these factors before he accepts and uses published validity coefficients. Our experience in the Services made it abundantly clear that all of these factors are exceedingly important in correlation studies. Dealing with so many different types of populations, with varied backgrounds of educational and vocational experience and selected on many different bases, it was possible to see how significant these factors were in the interpretation of validity coefficients.

A third factor which will greatly influence the magnitude of a validity coefficient is the nature of the criterion with which the test is correlated. Unfortunately many test builders have indiscriminately accepted whatever criterion of performance was available and have correlated their test scores with this criterion as a matter of expediency. The consequence has been the publication of many misleading coefficients of validity which the guidance worker and counselor have often unwittingly accepted and used in their work. The net result of this practice has been a complete misinterpretation of test scores and loss of faith in tests on the part of both counselors and clients.

The types of criterion measures which are most frequently used in educational and industrial personnel work are: (1) grades in training courses, (2) ratings of performance and (3)

objective outcomes such as sales per month, number of words typed per minute, frequency of errors or amount of work spoiled. Each of these criteria has advantages and disadvantages which should be recognized in the interpretation of validity coefficients.

The criterion measure which is most familiar to educational personnel workers is the school grade. The important question for us is: Does the school grade have a uniform interpretation from subject to subject, from class to class, and from school to school? The answer is probably "No." In some classes the grade in the course may be determined largely by the number of class recitations, by punctuality, attentiveness and neatness of work. In another course, or class within the same department, the major emphasis may be upon the answers to objective-type quizzes, with no weight given to any of the factors mentioned earlier. In some laboratory courses considerable weight may be given to laboratory technique; in others, laboratory work may be largely regarded as a necessary chore. Each of the criteria described may be quite reliable but this fact provides no assurance that the criteria are equally appropriate. Certainly they do not include the same things and aptitude tests which are correlated with these different criteria will show very different validity coefficients, depending upon the particular criterion which is being considered.

Ratings made of school work or performance on the job are subject to many of the same defects mentioned above. Judges may vary greatly in their definitions of success and hence may actually employ entirely different yardsticks; one judge, for example, may emphasize greatly the personal qualities of the individual while another may consider only the actual products of the worker's efforts. And with respect to products turned out one judge may emphasize speed and another, quality. Needless to say an aptitude test which will predict success as appraised by one judge will not necessarily predict success as appraised by another judge. Unless the test user knows the exact nature of the criterion with which the test was correlated he will not be in a position to determine accurately the validity of the test.

Objective criteria such as the number of words typed per minute or the number of errors per typed page are often regarded as the most desirable types of criteria. Actually these criteria too may be inappropriate. It would hardly be logical, for example, to evaluate a receptionist-secretary on such a narrow basis. The number of words typed per minute is simply not a comprehensive enough criterion to evaluate so complex a job. A secretarial aptitude test correlated with such a narrow criterion would, therefore, have a validity coefficient which would be quite inappropriate for use in the selection of a receptionist-secretary. While objective criteria are generally more reliable than subjective criteria there is no assurance that they will always be more appropriate in a particular situation. There are many instances in the literature where so-called special aptitude tests have produced entirely misleading results because of the narrowness of the criterion employed.

During the war years military psychologists had many opportunities to observe the effect of different types of criterion measures upon the validity coefficients of aptitude tests. Three illustrations of such experience will be presented here.

The first illustration which I should like to present is the effect upon validity coefficients of the introduction of an objective, written achievement test in an officer training program in the U. S. Navy. Before the introduction of the achievement test the students in this school had been evaluated on the basis of weekly grades and ratings in a "laboratory" situation. The ratings were felt to be the most important type of evaluation and hence they were allowed to constitute two-thirds of the final grade, that is, the ratings were multiplied by two and added to the average weekly quiz grades in computing the student's final grade.

The correlations between aptitude test scores and this criterion of school success were, without exception, low. For example, the correlation of the verbal part of the *Officer Classification Test* with the final grade was .18; the correlation between the mathematical part of this test and the final grade was .01. After the introduction of the final achievement examination and the elimination of the ratings from the criterion these corre-



lations increased to .32 and .49, respectively. While one may not conclude from these data that the new criterion was necessarily better than the old one, it can be said that the new criterion was predictable by means of our aptitude tests, whereas the old one was not. It should be added that in the judgment of the school staff the new criterion constituted a great improvement over the one which had formerly been used.

A second illustration is furnished by the changes in the validity of aptitude tests introduced by the use of standardized written achievement and performance tests in the evaluation of trainees in a basic engineering school. Before the introduction of performance testing it was customary for the school to use informal written tests extensively and while shop ratings were used, their effective weight was quite small. In these circumstances the *Arithmetical Reasoning Test* of the *Basic Test Battery* was the best predictor of success in this school; after the introduction of the standardized written achievement and performance tests and the resultant greater emphasis on practical ability, the *Mechanical Knowledge Test* became the best predictor and the verbal and numerical aptitudes became less important and less appropriate.

A third illustration is drawn from our experience in attempting to correlate our predictive measures with success aboard ship. Prior to the inauguration of the shipboard studies the predictor variables had been correlated with success in school. In these circumstances the best single predictor of success for radiomen was a code learning aptitude test. The *General Classification Test* (a test of general mental ability) correlated quite low with success in school (.26). When the correlations between predictor variables and success aboard ship were computed it was found that the highest correlation was between the *General Classification Test* and the criterion (.38). Evidently the criterion of shipboard performance placed the major emphasis upon all-around versatility whereas success in radio school was dependent primarily upon the ability to learn the code. In counseling work we often assume that an aptitude test which correlates with school success will also correlate with success on the job. Data such as these indicate that this may not always be the case.

In summary, it has been the experience of military psychologists that the so-called validity of a test will be greatly influenced by the nature of the criterion. Many writers in our professional literature have pointed out the same facts. Unfortunately such data are not displayed too prominently in the test manuals accompanying many of the tests which we use in guidance and counseling. It is, therefore, the responsibility of all of us who construct tests to describe the nature of our criterion as well as to describe the experimental and standardization populations; as test users it should increasingly be our responsibility to demand such information and to use it intelligently whenever it is made available to us.

•

RELATIONSHIP BETWEEN MILITARY OCCU-  
PATIONAL SPECIALTY AND ARMY  
GENERAL CLASSIFICATION  
TEST STANDARD  
SCORE<sup>1</sup>

NAOMI STEWART

Cooperative Test Service of the American Council on Education

THIS paper presents findings concerning the relationship between *Military Occupational Specialty* and *Army General Classification Test* Standard Score for 68,325 white and colored enlisted men in 220 military occupational specialties. The basic data for the study were obtained from War Department Machine Records Survey No. 4, taken September 30, 1944.

These War Department surveys were taken every three months and yielded data for approximately 2 per cent of all Army personnel—or roughly 150,000 individuals—on variables such as A.G.C.T. standard score, years of education, age, main civilian occupation, military occupational specialty, present duty assignment, etc. In War Department Machine Records Survey No. 4, data of this nature were obtained for all Army personnel whose serial numbers ended either in 19 or 75.

The *Army General Classification Test* has already been so thoroughly described (5, 6), that no additional elaboration is necessary. *Military Occupational Specialty* (M.O.S.) is, how-

---

<sup>1</sup> This paper is based in part on data which the writer analyzed while employed by the Personnel Research Section of the Adjutant General's Office, reported as Personnel Research Study No. 675, Part III (7). The analysis was made under the direction of Drs. E. A. Rundquist and Reign H. Bittner, whose help throughout the development of PRS Study No. 675 is here gratefully acknowledged. For the organization of the Personnel Research Section at the time the study was done, see (4). The study was initiated at the suggestion of Dr. Walter V. Bingham, Chairman of the Committee on Military Personnel Advisory to the Adjutant General, who is also to be credited with valuable suggestions, as are Drs. E. Donald Sisson, Erwin K. Taylor, Edwin R. Henry and Marion W. Richardson.

ever, a term which could well be amplified. Contrary to popular impression, a man was not once and for all classified in a particular specialty. Successful completion of a certain course of training, increased experience in a given assignment, etc., would result in a change of classification. Furthermore, a man's duty assignment was not by any means always the same as his M.O.S. The present study is based on M.O.S. classifications as of September 30, 1944, for the 103,998 white and colored enlisted men in Survey 4. Since A.G.C.T. scores were not available for officers, they were not included in the study.

### *Procedure and Results*

It was not possible to utilize data for 18,791 enlisted men because of obvious errors and ambiguities which were found in some of the variables under consideration, resulting in impossible A.G.C.T. scores, or non-existent M.O.S.'s, or disproportionately large numbers of cases in rare or infrequent M.O.S.'s, etc. War Department Machine Records data are especially valuable because of the effort that was made to obtain a *complete* 2 per cent sample of the Army, but the necessity for obtaining and punching data wherever our forces were stationed, despite what were sometimes extremely adverse conditions, resulted inevitably in errors of the type mentioned.

For the remaining 85,207 enlisted men, distributions of A.G.C.T. scores were obtained, broken down by M.O.S. and also by race within each M.O.S. For all M.O.S.'s in use on September 30, 1944, for which data on at least 25 cases were available, A.G.C.T. medians were obtained. For those M.O.S.'s for which data on at least 50 cases were available, the 10th, 25th, 75th and 90th percentiles of the A.G.C.T. distributions were also obtained.

Table 1 presents this information for 68,325 enlisted men in 220 military occupational specialties. Data for the remaining 16,882 cases analyzed are not given, either because the N's in the M.O.S. groups in which these men were coded were smaller than 25, or because their M.O.S.'s were no longer in use. Military Occupational Specialties are listed in Table 1 in descending order of A.G.C.T. median.

TABLE 1

*Selected Percentile Points for A.G.C.T. Distributions of White and Colored Enlisted Men in Military Occupational Specialties Listed in TM 12-427 as Current in September, '44 (M.O.S.'s in Descending Order of Median A.G.C.T. Score; Only M.O.S.'s With at Least 25 Cases Included, Based on Data from Machine Records Survey #4, Taken 30 September 1944)*

M.O.S.	R*	P <sub>90</sub>	P <sub>75</sub>	P <sub>50</sub>	P <sub>25</sub>	P <sub>10</sub>	No.	Q
787. Weather Forecaster	W	..	..	134	...	...	25	..
	T	...	..	134	...	...	26	..
784. Weather Observer	W	140	133	127	120	114	151	6.5
	T	140	134	127	120	114	157	7.0
275. Classification Specialist	W	142	133	124	117	110	229	8.0
	T	142	133	124	117	109	235	8.0
805. Cryptographic Technician	W	140	132	124	114	105	178	9.0
	T	140	131	124	113	103	184	9.0
683. Bombsight Mechanic	W	141	134	124	118	111	76	8.0
	T	142	134	124	118	111	79	8.0
953. Radar Repairman, Reporting Equipment	W	139	132	124	118	109	78	7.0
	T	139	132	124	118	109	78	7.0
187. Repeaterman, Telephone	W	140	137	125	119	110	54	9.0
	T	140	136	124	115	108	60	10.5
772. Liaison Pilot	W	..	...	123	...	...	31	...
	T	..	...	123	...	...	31	...
938. AAF Gunnery Instructor	W	134	128	122	113	106	141	7.5
	T	134	128	122	114	106	142	7.0
301. Investigator	W	141	133	122	114	108	130	9.5
	T	141	131	122	114	108	134	8.5
622. Finance Technical Clerk	W	140	132	122	116	110	118	8.0
	T	140	132	122	116	110	121	8.0
274. Writer, Military Subjects	W	137	130	122	117	110	59	6.5
	T	137	130	122	117	110	59	6.5
649. Radio Repairman, Fixed Station	W	...	..	122	...	..	36	..
	T	...	...	122	...	..	36	...
095. Central Office Repairman	W	...	...	122	...	..	30	..
	T	..	..	122	..	..	30	...
502. Administrative NCO	W	137	129	121	114	107	930	7.5
	C	..	..	110	..	..	32	...
	T	137	129	121	114	106	973	7.5
624. Finance Clerk	W	139	133	122	115	109	77	9.0
	T	139	133	121	115	109	79	9.0
149. Pharmacist	W	...	..	120	...	...	35	...
	T	...	..	121	...	...	39	...
858. Medical Laboratory Technician	W	137	128	120	111	102	150	8.5
	T	137	128	120	111	99	163	8.5
629. Student	W	140	131	120	113	105	119	9.0
	T	140	131	120	111	102	133	10.0
623. Finance Typist Clerk	W	139	129	120	113	108	86	8.0
	T	139	129	120	113	108	89	8.0
658. Link Trainer Instructor	W	138	129	120	111	107	79	9.0
	T	138	129	120	111	107	79	9.0

\* R—Race: W—White; C—Colored, U—Unknown, T—Total. Total figures are for white, colored and unknown.

TABLE 1 (Continued)

M.O.S.	R*	P <sub>20</sub>	P <sub>50</sub>	P <sub>60</sub>	P <sub>70</sub>	P <sub>80</sub>	No.	Q
647. Radio Repairman, Aircraft Equipment	W	140	132	122	115	108	73	8.5
	T	140	131	120	114	107	75	8.5
719. Oxygen Plant Operator	W	...	...	120	...	...	28	...
	T	...	...	120	...	...	28	...
213. Stenographer	W	138	128	119	112	104	177	8.0
	T	138	128	119	112	103	185	8.0
862. Radar Mechanic (IFF)	W	132	127	119	111	105	98	8.0
	T	135	127	119	112	105	99	7.5
076. Draftsman, Topographic	W	134	126	119	110	104	76	8.0
	T	134	124	119	109	103	78	7.5
952. Radar Repairman Gun-Laying Equipment (Designated Set)	W	139	129	119	112	106	56	8.5
	T	139	129	119	111	106	58	9.0
777. Radio Operator, High Speed, Automatic	W	...	...	119	...	...	46	...
	T	...	...	119	...	...	47	...
955. Radar Repairman, Airborne Equipment (Designated Set)	W	...	...	119	...	...	39	...
	T	...	...	119	...	...	43	...
925. Aircraft Engineering Technician	W	...	...	120	...	...	36	...
	T	...	...	119	...	...	38	...
867. Radar Mechanic, Bombardment	W	...	...	119	...	...	32	...
	T	...	...	119	...	...	32	...
648. Radio Repairman	W	138	128	118	110	102	239	9.0
	T	138	128	118	109	101	247	9.5
631. Intelligence NCO	W	138	128	119	111	102	138	8.5
	T	136	128	118	110	99	152	9.0
239. Teletype Mechanic	W	134	124	118	107	101	96	8.5
	T	134	124	118	107	101	97	8.5
814. Operations NCO	W	135	127	119	111	97	85	8.0
	T	135	126	118	110	95	91	8.0
070. Draftsman	W	...	...	118	...	...	37	...
	T	...	...	118	...	...	38	...
004. Aerial Phototopographer	W	...	...	121	...	...	36	...
	T	...	...	118	...	...	37	...
405. Clerk-Typist	W	134	127	117	108	99	2550	9.5
	C	120	113	103	91	80	138	11.0
	T	134	126	117	108	97	2705	9.0
678. Power Turret and Gun-sight Mechanic	W	134	125	117	108	102	100	8.5
	T	134	125	117	108	102	101	8.5
859. Pharmacy Technician	W	131	126	119	108	95	50	9.0
	T	131	125	117	108	93	51	8.5
659. Instructor (Designated Subject)	W	137	126	118	102	88	165	12.0
	T	137	125	116	101	88	184	12.0
442. Entertainment Specialist	W	...	...	117	...	...	46	...
	T	128	122	116	108	91	50	7.0
438. Bandsman, French Horn	W	...	...	118	...	...	33	...
	T	...	...	116	...	...	36	...
754. Radio Mechanic, AAF	W	132	124	115	107	100	443	8.5
	T	132	124	115	107	100	448	8.5
757. Radio Operator-Mechanic-Gunner, AAF	W	132	124	115	109	104	337	7.5
	T	132	124	115	109	103	340	7.5

TABLE 1 (Continued)

M.O.S.	R*	P <sub>90</sub>	P <sub>75</sub>	P <sub>50</sub>	P <sub>25</sub>	P <sub>10</sub>	No	Q
756. Radio Operator, AAF	W	131	123	115	107	99	281	8.0
	T	131	123	115	107	99	289	8.0
667. Message Center Clerk	W	132	125	116	105	94	238	10.0
	T	132	124	115	105	92	247	9.5
542. Communications Chief	W	129	124	115	105	94	220	9.5
	T	130	124	115	105	93	230	9.5
228. Survey and Instrument Man	W	136	126	115	107	95	125	9.5
	T	136	126	115	106	94	130	10.0
097. Installer-Repairman, Telephone and Telegraph	W	128	123	115	99	91	130	12.0
	T	128	123	115	99	91	130	12.0
283. Athletic Instructor	W	134	124	116	109	96	121	7.5
	T	133	124	115	106	95	129	9.0
759. Radio Operator, CNS	W	129	123	115	108	102	124	7.5
	T	129	122	115	108	101	127	7.0
740. Radio Operator, Intermediate Speed	W	128	124	117	104	99	56	10.0
	T	128	124	115	104	94	58	10.0
825. Medical Supply NCO	W	...	...	115	...	.	47	.
	T	...	...	115	...	..	47	..
261. Wire Chief, Telephone and Telegraph	W	...	...	116	...	.	44	...
	T	...	...	115	...	...	45	...
322. Refrigeration Mechanic	W	...	...	115	...	..	38	...
	T	...	...	115	...	...	38	...
769. Chief Storekeeper, Railway	W	...	..	115	...	..	30	..
	T	...	..	115	...	...	30	...
821. Quartermaster Supply Technician	W	132	124	115	104	94	791	10.0
	C	121	114	99	86	73	98	14.0
	T	131	123	114	101	91	892	11.0
766. Radio Operator, High Speed, Manual	W	133	123	114	106	97	316	8.5
	T	133	123	114	106	96	324	8.5
237. Teletype Operator	W	131	122	114	103	94	236	9.5
	T	131	122	114	103	92	244	9.5
686. Airplane Instrument Mechanic	W	132	124	114	107	96	181	8.5
	T	132	124	114	107	96	186	8.5
945. Photographic Laboratory Technician	W	132	124	114	105	95	171	9.5
	T	132	124	114	105	95	175	9.5
687. Airplane Propeller Mechanic	W	127	122	115	106	98	142	8.0
	T	127	122	114	105	98	149	8.5
750. Airplane Maintenance Technician	W	128	124	114	107	100	101	8.5
	T	128	124	114	107	100	102	8.5
432. Bandsman, Clarinet	W	130	126	116	106	97	54	10.0
	T	130	124	114	105	93	62	9.5
152. Photographer	W	...	...	117	...	...	34	...
	T	...	...	114	...	...	36	..
727. Water Supply Technician	W	...	..	118	...	..	27	...
	T	...	...	114	...	...	31	...
518. Ground Aircraft Observer	W	...	...	114	...	...	27	...
	T	...	..	114	..	...	31	..
826. AAF Supply Technician	W	130	122	114	103	94	257	9.5
	T	130	122	113	104	93	263	9.0

TABLE I (Continued)

M.O.S.	R*	P <sub>90</sub>	P <sub>75</sub>	P <sub>50</sub>	P <sub>25</sub>	P <sub>10</sub>	No.	Q
264. X-Ray Technician	W	131	124	115	106	92	102	9.0
	T	131	124	113	104	91	107	10.0
674. Message Center Chief	W	137	123	114	107	100	73	8.0
	T	137	121	113	106	99	79	7.5
433. Bandsman, Cornet or Trumpet	W	127	121	114	106	99	68	7.5
	T	127	121	113	104	98	73	8.5
552. Control Tower Operator	W	134	126	113	103	95	66	11.5
	T	134	124	113	104	95	67	10.0
941. Camera Technician	W	...	...	113	...	...	41	...
	T	...	...	113	...	...	47	...
791. Air Operations Specialist	W	...	...	112	...	...	41	...
	T	...	...	113	...	...	42	...
804. Camouflage Technician	W	...	...	114	...	...	35	...
	T	...	...	113	...	...	40	...
738. Intercept Operator, G	W	...	...	113	...	...	35	...
	T	...	...	113	...	...	35	...
039. Cable Splicer, Telephone and Telegraph	W	...	...	113	...	...	30	...
	T	...	...	113	...	...	34	...
166. Powerman	W	...	...	113	...	...	27	...
	T	...	...	113	...	...	27	...
776. Radio Operator, Low Speed	W	129	122	113	102	91	886	10.0
	C	...	...	97	...	...	25	...
	T	129	122	112	102	91	914	10.0
911. Airplane Armorer	W	128	121	112	104	92	795	8.5
	T	128	120	112	103	91	822	8.5
748. Airplane Mechanic-Gunner	W	128	120	112	103	95	479	8.5
	T	128	119	112	103	95	481	8.0
612. Airplane Armorer-Gunner	W	128	121	112	105	98	416	8.0
	T	128	121	112	105	98	426	8.0
645. Fire Control Instrument Operator, Field Artillery	W	132	123	112	101	82	318	11.0
	T	132	123	112	97	74	345	13.0
528. Airplane Hydraulic Mechanic	W	...	...	112	...	...	47	...
	T	...	...	112	...	...	47	...
236. Telegraph Operator	W	...	...	112	...	...	29	...
	T	...	...	112	...	...	29	...
056. Postal Clerk	W	128	122	112	100	86	571	11.0
	C	...	...	84	...	...	43	...
	T	128	121	111	97	81	617	12.0
673. Medical NCO	W	128	121	111	99	89	289	11.0
	T	128	121	111	99	89	308	11.0
684. Airplane Power Plant Mechanic	W	121	117	111	105	94	58	6.0
	T	121	117	111	105	94	59	6.0
227. Surveyor	W	...	...	116	...	...	25	...
	T	...	...	111	...	...	34	...
966. Mechanic, Automotive, Track Vehicle (Third Echelon)	W	...	...	112	...	...	27	...
	T	...	...	111	...	...	29	...
435. Bandsman, Snare Drum	T	...	...	111	...	...	25	...
747. Airplane and Engine Mechanic	W	126	119	110	100	89	2729	9.5
	C	...	...	95	...	...	48	...
	T	126	119	110	100	88	2793	9.5



TABLE 1 (Continued)

M.O.S.	R*	P <sub>100</sub>	P <sub>75</sub>	P <sub>50</sub>	P <sub>25</sub>	P <sub>10</sub>	No.	Q
685. Airplane Electrical Mechanic	W	127	120	110	104	90	189	8.0
	T	126	120	110	103	90	194	8.5
922. Instrument Repairman, Fire Control	W	128	120	110	101	93	53	9.5
	T	128	121	110	102	93	55	9.5
617. Altitude Chamber Technician	W	...	..	110	..	...	37	...
	T	...	..	110	...	...	37	...
441. Bandsman, Tuba	W	...	...	111	...	...	27	...
	T	..	...	110	...	..	29	..
815. Ordnance Supply NCO	W	..	...	113	...	..	26	..
	T	...	...	110	..	...	27	..
114. Machinist	W	126	119	109	97	79	292	11.0
	T	126	119	109	97	78	302	11.0
761. Scout	W	131	124	111	97	78	245	13.5
	T	130	122	109	86	71	268	18.0
903. Small Arms Weapons Mechanic	W	127	118	109	99	90	72	9.5
	T	127	118	109	99	85	78	9.5
915. Artillery Mechanic, Heavy Antiaircraft	W	...	...	109	...	..	46	...
	T	..	..	109	..	..	46	...
167. Lithographic Pressman	W	...	...	109	...	...	25	..
	T	..	...	109	...	...	25	..
861. Surgical Technician	W	127	119	108	95	81	971	12.0
	T	127	119	108	94	79	1005	12.5
555. Airplane Sheet Metal Worker	W	124	117	108	99	86	336	9.0
	T	124	117	108	98	87	347	9.5
348. Parts Clerk, Automotive	W	128	119	109	97	83	259	11.0
	T	127	119	108	96	82	271	11.5
762. Airplane Engine Repairman	W	126	119	108	101	89	127	9.0
	T	126	119	108	100	89	130	9.5
802. Artillery Mechanic, Minor Maintenance	W	126	118	108	96	83	111	11.0
	T	124	117	108	95	78	121	11.0
527. Antiaircraft Range Section NCO	W	132	118	110	95	82	112	11.5
	T	132	118	108	94	77	118	12.0
573. Welder, Aircraft	W	123	115	108	98	88	81	8.5
	T	123	115	108	98	87	83	8.5
337. Foreman, Automotive Repair Shop	W	124	116	109	93	85	53	11.5
	T	124	116	108	94	85	55	11.0
692. Height Finder Observer	W	...	..	108	...	...	35	..
	T	...	...	108	..	...	38	...
812. Heavy Weapons NCO	W	...	...	108	...	...	34	...
	T	...	..	108	..	...	36	..
118. Small Boat Operator	W	...	...	108	...	...	29	..
	T	...	...	108	...	...	29	...
724. Range Section Operator, Coast Artillery	W	...	...	108	...	..	27	...
	T	...	...	108	...	...	29	...
510. Information Center Operator	T	...	...	108	...	...	26	...
835. Supply Clerk	W	126	119	108	99	80	1106	10.0
	C	115	106	92	82	73	52	12.0
	T	125	117	107	93	79	1170	12.0
514. Radar Crewman (Designated Set)	W	128	119	108	93	80	403	13.0
	T	128	119	107	93	80	423	13.0

TABLE 1 (Continued)

M.O.S.	R*	P <sub>0</sub>	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	No.	Q
559. Glider Mechanic	W	...	...	107	...	...	49	...
	T	...	...	107	...	...	50	...
912. Electrician, Automotive	W	...	...	108	...	...	28	...
	T	...	...	107	...	...	29	...
282. Office Machine Service- man	W	...	...	107	...	...	28	...
	T	...	...	107	...	...	28	...
191. Rodman and Chainman, Surveying	W	...	...	107	...	...	25	...
	T	...	...	107	...	...	26	...
511. Armorer	W	126	116	106	93	76	374	11.5
	T	125	114	106	90	75	399	12.0
413. Motor Inspector	W	...	...	106	...	...	42	...
	T	...	...	106	...	...	42	...
855. Dental Technician	W	127	119	105	90	74	108	14.5
	T	127	119	105	89	73	113	15.0
067. Dental Laboratory Tech- nician	W	125	118	106	98	87	104	10.0
	T	125	118	105	97	86	110	10.5
610. Antitank Gun Crewman	W	128	118	107	89	74	102	14.5
	T	128	118	105	89	74	106	14.5
660. Tank Mechanic, Minor Maintenance	W	...	...	105	...	...	31	...
	T	...	...	105	...	...	31	...
078. Electrician	W	125	117	108	93	77	271	12.0
	C	...	...	69	...	...	27	...
	T	124	115	104	86	72	299	14.5
578. Observation Station Operator, Coast Artillery	W	121	116	104	88	72	55	14.0
	T	121	116	104	88	74	57	14.0
946. Searchlight NCO	W	129	115	105	95	85	52	10.0
	T	126	114	104	94	82	55	10.0
145. Painter, Sign	T	...	...	104	...	...	25	...
059. Foreman, Construction	W	129	120	109	95	81	444	12.5
	C	99	89	71	59	48	114	15.0
	T	127	116	103	83	65	570	16.5
252. Foreman, Warehouse	W	125	115	103	89	76	181	13.0
	T	125	115	103	86	72	194	14.5
309. Telephone Operator	W	124	116	103	93	84	104	11.5
	T	124	116	103	93	80	110	11.5
650. Telephone Switchboard Operator	W	123	116	105	94	76	89	11.0
	T	122	115	103	92	74	98	11.5
137. Projectionist, Motion Picture	W	127	116	104	89	80	89	13.5
	T	127	116	103	89	80	91	13.5
913. Artillery Mechanic, Light	W	125	119	105	96	86	89	11.5
	T	125	119	103	96	86	90	11.5
033. Brakeman, Railway	W	120	111	103	89	74	51	11.0
	T	120	111	103	89	74	51	11.0
566. Duty NCO	W	128	116	105	90	75	474	13.0
	C	108	92	76	63	57	130	14.5
	T	127	114	101	81	67	609	16.5
013. Diesel Mechanic	W	120	113	102	92	83	58	10.5
	T	120	113	101	89	82	65	12.0
409. Medical Technician	W	124	114	101	88	76	883	13.0
	C	...	...	81	...	...	40	...
	T	124	113	100	87	74	929	13.0

TABLE 1 (Continued)

M.O.S.	R*	P <sub>90</sub>	P <sub>75</sub>	P <sub>50</sub>	P <sub>25</sub>	P <sub>10</sub>	No.	Q
824. Mess Sergeant	W	123	115	102	90	76	555	12.5
	C	111	97	77	63	50	73	17.0
	T	122	114	100	86	72	641	14.0
965. Mechanic, Automotive Wheel Vehicle (Third Echelon)	W	120	113	100	86	75	364	13.5
	T	120	113	100	86	75	376	13.5
803. Bugler	W	126	113	102	88	72	168	12.5
	T	125	111	100	86	66	194	12.5
196. Sanitary Technician	W	128	120	109	95	73	75	12.5
	C	..	..	77	..	..	33	..
	T	126	115	100	78	61	108	18.5
548. Fabric and Dope Mechanic	W	..	..	99	..	..	39	..
	T	..	..	100	..	..	41	..
521. Basic	W	127	118	104	85	69	7416	16.5
	C	99	79	65	57	44	1149	11.0
	U	..	..	72	..	..	32	..
	T	126	116	99	77	62	8597	19.5
641. Field Lineman	W	121	114	100	86	72	770	14.0
	C	..	..	71	..	..	27	..
	T	120	113	99	93	71	804	10.0
505. Ammunition NCO	W	125	116	106	92	80	213	12.0
	C	..	..	71	..	..	44	..
	T	124	114	99	83	68	263	15.5
256. Welder, Combination	W	123	112	101	90	77	224	11.0
	T	120	111	99	89	72	250	11.0
729. Pioneer	W	128	117	101	88	67	213	14.5
	T	127	116	99	82	63	229	17.0
037. Meat Cutter	W	117	110	99	85	68	96	12.5
	T	117	110	99	82	68	98	14.0
809. Decontaminating Equip- ment Operator	W	131	117	104	92	81	64	12.5
	T	124	114	99	85	66	79	14.5
822. Utilities NCO	W	..	..	100	..	..	49	..
	T	125	115	99	83	72	53	16.0
014. Automotive Mechanic (Second Echelon)	W	120	111	99	88	75	1973	11.5
	C	104	90	75	62	51	197	14.0
	T	120	110	98	85	70	2181	12.5
677. Military Policeman	W	120	111	99	82	70	1986	14.5
	C	99	90	74	64	56	94	13.0
	U	..	..	89	..	..	49	..
	T	120	110	98	80	69	2129	15.0
017. Baker	W	123	111	100	86	70	252	12.5
	C	..	..	74	..	..	35	..
	T	122	110	98	81	66	291	14.5
373. Sales Clerk	W	..	..	101	..	..	34	..
	T	..	..	98	..	..	40	..
533. Demolition Specialist	W	125	114	100	86	69	146	14.0
	T	123	113	97	80	64	169	16.5
081. Engineman, Operating	W	124	114	100	88	73	135	13.0
	C	..	..	67	..	..	28	..
	T	122	111	97	78	64	165	16.5
144. Painter, General	W	120	114	99	87	77	116	13.5
	T	119	113	97	83	71	129	15.0

TABLE 1 (Continued)

M.O.S.	R*	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>	No.	Q
786. Toxic Gas Handler	W	123	109	99	86	68	102	11.5
	T	123	109	97	85	66	110	12.0
319. Construction Equipment Mechanic	W	120	111	99	86	75	93	12.5
	T	120	111	97	84	73	101	13.5
356. Foreman, Labor	W	127	111	102	84	74	54	13.5
	T	122	109	97	77	65	78	16.0
605. Heavy Machine Gunner	W	123	113	97	83	71	441	15.0
	C	...	...	62	...	...	29	...
	T	122	111	96	80	65	474	15.5
844. Gun Crewman, Light Artillery	W	127	117	99	81	67	171	18.0
	T	126	117	96	79	66	179	19.0
608. Gun Crewman, Coast Artillery	W	117	109	96	77	70	89	16.0
	T	118	109	96	77	70	90	16.0
201. Sheet Metal Worker	W	124	113	101	89	79	50	12.0
	T	124	111	96	80	63	60	15.5
378. Motorecyclist	W	...	...	97	...	...	44	...
	T	116	112	96	52	73	51	30.0
607. Light Mortar Crewman	W	123	113	97	80	69	429	16.5
	T	122	112	95	79	67	451	16.5
601. Antiaircraft Artillery Automatic Weapons Crewman	W	119	109	96	82	70	353	13.5
	T	118	109	95	79	67	375	15.0
932. Special Vehicle Operator	W	115	106	96	83	73	288	11.5
	T	115	105	95	80	69	313	12.5
064. Power Shovel Operator	W	...	...	97	...	...	37	...
	T	...	...	95	...	...	49	...
745. Rifleman	W	124	112	95	78	66	3919	17.0
	C	99	85	71	61	48	260	12.0
	T	124	111	94	76	65	4193	17.5
604. Light Machine Gunner	W	122	109	94	77	66	398	16.0
	T	122	109	94	77	65	414	16.0
901. Munitions Worker	W	120	109	96	80	70	327	14.5
	C	...	...	68	...	...	27	...
	T	119	108	94	78	65	358	15.0
065. Seaman	W	120	111	94	74	62	115	18.5
	T	120	110	94	74	63	119	18.0
316. Automobile Serviceman	W	...	...	97	...	...	32	...
	T	...	...	94	...	...	41	...
839. Marine Engineman	W	...	...	94	...	...	26	...
	T	...	...	94	...	...	26	...
035. Carpenter, Heavy Construction	W	124	114	98	82	70	125	16.0
	T	122	110	93	76	63	149	17.0
164. Plumber	W	118	110	98	78	67	53	16.0
	T	118	109	93	76	67	58	16.5
204. Shoe Repairman	W	...	...	100	...	...	48	...
	T	116	110	93	75	60	56	17.5
864. Gun Crewman, Medium Artillery	W	...	...	93	...	...	33	...
	T	...	...	93	...	...	37	...
735. Full-Track Driver	W	...	...	95	...	...	33	...
	T	...	...	93	...	...	35	...

TABLE 1 (Continued)

M.O.S.	R*	P <sub>100</sub>	P <sub>75</sub>	P <sub>50</sub>	P <sub>25</sub>	P <sub>10</sub>	N <sub>0</sub>	Q
238. Lineman, Telephone and Telegraph	W	120	108	96	81	70	429	13.5
	C	99	86	76	62	54	99	12.0
	T	118	106	92	77	65	533	14.5
189. Rigger	W	121	110	93	81	68	280	14.5
	C	..	..	66	..	..	39	..
	T	120	109	92	75	63	324	17.0
383. Fire Fighter	W	120	108	95	80	65	147	14.0
	T	120	108	92	75	61	158	16.5
763. Searchlight Crewman	W	119	107	94	78	67	107	14.5
	T	119	106	92	78	65	123	14.0
845. Gun Crewman, Heavy Artillery	W	119	108	92	83	68	82	12.5
	T	119	107	92	83	68	86	12.0
121. Utility Repairman	W	119	108	93	79	69	343	14.5
	C	103	87	72	51	42	52	18.0
	T	118	106	91	76	62	402	15.0
050. Carpenter, General	W	120	110	98	82	68	202	14.0
	C	..	..	69	..	..	40	..
	T	119	108	91	75	62	248	16.5
359. Construction Machine Operator	W	120	110	96	80	68	197	15.0
	C	..	..	73	..	..	42	..
	T	119	109	91	74	63	240	17.5
846. Portable Power Generator Operator	W	119	108	92	76	64	187	16.0
	T	117	107	91	74	64	198	16.5
242. Toolroom Keeper	W	124	114	97	82	71	103	16.0
	C	..	..	66	..	..	28	..
	T	123	111	91	74	59	134	18.5
733. Reconnaissance Car Crewman	W	115	106	91	78	69	118	14.0
	T	115	106	91	78	68	119	14.0
194. Salvage Technician	W	119	111	104	89	81	54	11.0
	T	116	108	91	74	59	75	17.0
024. Blacksmith	W	127	107	95	81	70	51	13.0
	T	118	100	91	73	58	66	13.5
234. Tailor	T	..	..	91	..	..	28	..
746. Automatic Rifleman	W	124	113	93	73	60	497	20.0
	C	..	..	67	..	..	34	..
	T	123	112	90	72	63	534	20.0
931. Truck Driver, Heavy	W	116	109	93	79	69	407	15.0
	C	88	74	65	56	44	75	9.0
	T	115	106	90	74	60	485	16.0
102. Foreman, Laundry	W	..	..	93	..	..	28	..
	T	..	..	90	..	..	38	..
345. Truck Driver, Light	W	117	107	93	79	67	5003	14.0
	C	96	80	67	58	46	1300	11.0
	U	114	102	76	65	58	58	18.5
	T	115	104	89	72	61	6361	16.0
657. Medical Aidman	W	119	106	89	74	62	711	16.0
	C	106	93	72	59	46	59	17.0
	T	118	104	89	72	61	771	16.0
063. Crane Operator	W	119	112	98	81	72	56	15.5
	T	119	108	89	74	65	74	17.0
522. Duty Soldier I	W	114	105	89	75	65	630	15.0
	C	105	88	69	57	45	75	15.5
	T	114	103	88	73	61	710	15.0

TABLE 1 (Continued)

M.O.S.	R*	P <sub>90</sub>	P <sub>75</sub>	P <sub>50</sub>	P <sub>25</sub>	P <sub>10</sub>	No.	Q
084. Stationary Fireman	W	115	106	92	80	67	56	13.0
	T	113	100	88	68	57	76	16.0
529. Wrecker Crewman	W	115	107	89	76	63	53	15.5
	T	113	105	88	76	63	55	14.5
732. Amphibian Tractor Driver	W	...	...	88	...	...	30	..
	T	...	...	88	...	...	30	..
034. Bricklayer	T	...	...	88	...	..	26	...
014. Canvas Cover Repairman	T	...	...	88	..	...	25	...
060. Cook	W	116	105	90	66	41	723	19.5
	C	86	78	66	57	46	105	10.5
	U	...	...	74	...	...	33	...
	T	114	103	87	71	61	861	16.0
022. Barber	W	...	...	82	...	..	36	...
	T	..	...	82	...	...	38	...
103. Laundry Machine Operator	W	110	101	87	72	60	105	14.5
	C	...	...	67	...	...	43	...
	T	109	96	81	65	57	148	15.5
188. Duty Soldier II	W	111	107	93	76	69	84	15.5
	C	..	...	62	...	...	48	..
	T	110	100	80	65	46	132	17.5
199. Section Hand, Railway	W	113	102	87	68	62	60	17.0
	T	110	92	79	63	52	78	14.5
473. Winch Operator	W	...	..	92	...	...	33	...
	C	95	79	69	60	50	67	9.5
	T	108	89	74	62	50	116	13.5
934. Amphibian Truck Driver	W	...	...	86	..	...	31	...
	C	...	...	65	...	...	27	...
	T	106	92	74	64	60	64	14.0
590. Duty Soldier III	W	111	99	84	71	60	418	14.0
	C	89	77	63	52	42	339	12.5
	U	...	...	63	...	...	30	...
	T	105	90	73	59	46	787	15.5
271. Longshoreman	W	117	106	90	78	67	244	14.0
	C	89	77	65	56	46	448	10.5
	U	...	...	70	...	...	55	...
	T	106	89	72	60	50	747	14.5

It is obvious from Table 1 that there is a definite A.G.C.T. hierarchy for the 220 M.O.S.'s studied, with medians ranging from 134 for Weather Forecaster to 72 for Longshoreman, based on data for white and colored enlisted men. The 10th percentile for twelve of the M.O.S.'s studied is higher than the 90th percentile for Longshoreman (106): Weather Observer; Classification Specialist; Bombsight Mechanic; Radar Repairman, reporting equipment; Repeaterman, Telephone; Investigator; Finance Technical Clerk; Writer, military subjects; Finance

Clerk; Finance Typist Clerk; Link Trainer Instructor; Radio Repairman, aircraft equipment. The M.O.S. with the highest  $P_{10}$  (114) is Weather Observer. There are nine M.O.S.'s with  $P_{90}$ 's lower than this: Stationary Fireman; Wrecker Crewman; Laundry Machine Operator; Duty Soldier II; Railway Section Hand; Winch Operator; Amphibian Truck Driver; Duty Soldier III; Longshoreman. The A.G.C.T. medians for colored personnel are uniformly lower than A.G.C.T. medians for white personnel in the same M.O.S.'s.

In interpreting the results given in Table 1 it is necessary to remember that these values reflect all of the factors of previous selection and all of the accidents of assignment. Sisson (3) gives an account of some of the factors operative.

#### *Comparison With Previous Findings*

In a previous study of the relationship between A.G.C.T. standard score and main civilian occupation (8), the results revealed an interesting negative association between the position of the civilian occupations in the A.G.C.T. hierarchy and the semi-interquartile range ( $r = -.80$ ). When distributions of scores corresponding to the 10th, 25th, 50th, 75th and 90th percentiles of the A.G.C.T. distributions for the various civilian occupations studied were compiled, the variability of the  $P_{90}$ 's was found to be about half of the variability of the  $P_{10}$ 's, with the variability decreasing systematically as one went from the distribution of  $P_{10}$ 's through the distributions of  $P_{25}$ 's,  $P_{50}$ 's,  $P_{75}$ 's, and  $P_{90}$ 's.

A parallel analysis was made of the M.O.S. data to determine whether these relationships also hold for the A.G.C.T. distributions of the various military occupational specialties.

A plot of A.G.C.T. median against  $Q$  for the total group in each M.O.S. where both median and  $Q$  are available reveals again a very highly negative relationship, with an  $r$  of about  $-.80$ . This indicates that the lower the military occupational specialty in the A.G.C.T. hierarchy, the greater the variability of the individuals in that specialty with respect to the A.G.C.T. score.

Table 2 gives means and sigmas for the distributions of

A.G.C.T. scores corresponding to the 10th, 25th, 50th, 75th, and 90th percentiles in the total distributions for the M.O.S.'s listed in Table 1.

The results indicate that military occupational specialties, like civilian occupational specialties, are more variable with respect to their poorest than with respect to their best representatives, when poorest and best are defined in terms of A.G.C.T. scores. The variability among the 10 per cent with the lowest A.G.C.T. scores in the 159 occupations included is more than twice as great as the variability among the 10 per cent with the highest A.G.C.T. scores.

A study by Harrell (1) gives information on the A.G.C.T. distributions of about 200 M.O.S.'s, based on data for all AAF enlisted men in the continental United States in August, 1943.

TABLE 2

*Mean and Standard Deviation for the Distribution of A.G.C.T. Scores Corresponding to the 10th, 25th, 50th, 75th, and 90th Percentiles for the Total Group in Each M.O.S.*

	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>
Number .....	159	159	220	159	159
Mean .....	79.2	92.0	105.7	116.7	126.0
Sigma .....	19.4	18.9	11.5	9.2	8.1

He omitted cases where records were incomplete and where there were fewer than 100 men in a specialty. Except for the fact that Harrell's study does not include personnel overseas, whereas Survey 4 figures do, the only major difference between the two sets of data lies in the fact that his results are only for AAF personnel and for some personnel attached to the AAF, while the present study gives figures for somewhat less than 2 per cent of all Army personnel (including AAF personnel whose serial numbers ended in 19 or 75). There were 135 M.O.S.'s common to both studies, for 110 of which *Q*'s were available in the present study. The significance of difference between the Harrell median and that obtained in the present study was tested for these 110 M.O.S.'s.

Since no *Q* values were given for Harrell's data, these were estimated as equal to the *Q*'s found for the respective M.O.S.'s in the present study. The probable error of each median was



approximated, using Table 6 in (8), and a critical ratio of 4 maintained as a standard in testing significance of difference between medians.

The comparison reveals 28 significant differences, of which 22 are in favor of the Army-wide median. Rifleman was higher by 19 A.G.C.T. score points in the present study; Basic by 16. The Army-wide data also give higher medians for Medical NCO, Medical Technician, and Surgical Technician.

Eight M.O.S.'s pertaining to communications were apparently filled by lower calibre personnel in the AAF than in the Army as a whole: Repeaterman, Telephone; Installer Repairman, telephone and telegraph; Message Center Clerk; Field Lineman; Lineman, telephone and telegraph; Radio Operator, low speed; Radio Operator, high speed, manual; and Teletype Operator. The A.G.C.T. median for Communications Chief, however, is significantly higher for AAF personnel.

Other M.O.S.'s with significantly higher medians for the Army-wide sample are: Quartermaster Supply Technician, Clerk-Typist, Armorer, Military Policeman, Automotive Parts Clerk, Automotive Mechanic (2nd echelon), Postal Clerk, Radar Repairman (reporting equipment) and Truck Driver, light.

The largest difference found in favor of the AAF median was that for Buglers (13 score points). The probability is extremely high that AAF buglers were more intelligent than buglers in the Army as a whole. AAF Ammunition NCO's and Riggers were also higher in A.G.C.T. median than those in the present study, by 7 score points; AAF Cooks by 3 score points.

One significant difference obtained is somewhat puzzling. Radio Operator, AAF, was significantly higher in Harrell's study than Radio Operator, AAF, in the present study, (C.R. 5.00). Since M.O.S.'s with AAF designations were presumably *exclusive* to the AAF, this may point to sampling contamination; it may be a function of the fact that the Gaussian curve is asymptotic to the base line; or indicate a true difference between AAF Radio Operators who were in the continental U.S. in 1943 and those both at home and abroad in 1944.

The findings reported in this section do not appear to substantiate the claim often made that AAF enlisted personnel were superior in A.G.C.T. score to those in other branches of the Army (e.g., 2, p. 455). Certain M.O.S.'s, however, were more critical in the Air Force than they were in the Army, and vice versa. In addition, there were more technical specialties in the Air Force requiring high skills. Both of those factors should be considered in interpreting the results obtained here.

A comparison of A.G.C.T. medians for men who had been coded as falling into certain civilian 'occupational specialty classifications (8) with the medians obtained in the present study for men in the same military occupational specialties was made for 40 occupations on which data for at least 50 cases were available in both studies. Only data for white enlisted men in the present study were used since (8) had included only white enlisted men.

The median for the M.O.S. was higher than that for the same *civilian* occupation in 5 instances, but none of the differences were significant. There were 12 significant differences in the 40 comparisons made and all 12 gave higher medians for the civilian than for the military groups.

The medians for Postal Clerk, Warehouse Foreman, Meat Cutter, Telephone and Telegraph Lineman, Utility Repairman, Toolroom Helper, Firefighter, and Cook were significantly lower (by six or more A.G.C.T. score points) in the military than in the civilian occupational specialties. Other significant differences giving higher medians for the civilian occupational groups were obtained for Clerk-Typist, Automotive Mechanic, Heavy Construction Carpenter and Truck Driver, light.

The evidence appears to indicate that the Army was able to utilize men of somewhat poorer calibre for these occupations than are found in these occupations in civilian life.

### *Summary and Conclusions*

Selected percentile points in the A.G.C.T. distributions for white and colored enlisted men in 220 Military Occupational Specialties have been presented, based on data for 68,325 men in War Department Machine Records Survey 4, representing approximately 2 per cent of the Army.

A marked A.G.C.T. hierarchy has been noted, with A.G.C.T. distributions increasing in variability as the medians decrease. More variability among M.O.S.'s has been found with respect to the poorest men in each of the M.O.S.'s studied than among the best men in each.

A.G.C.T. medians for colored personnel in the M.O.S.'s studied are uniformly lower than those for white.

A comparison of these results with those reported by Harrell for AAF personnel reveals that there were significant differences for 22 M.O.S.'s in favor of the Army-wide sample; for 6 in favor of the AAF.

A comparison of the M.O.S. data with data for the same civilian occupational specialties reveals that the calibre of men in most of the civilian occupations was higher than that of men in the same military occupational specialties.

#### REFERENCES

1. Harrell, T. W. "Army General Classification Test Results for Air Forces Specialists." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 341-349.
2. Harrell, T. W. "Applications of Psychology in the American Army." *Psychological Bulletin*, XLII (1945), 453-460.
3. Sisson, E. D. "Measuring Minds." *Army Information Digest*, (1947), 43-48.
4. Staff, Personnel Research Section, Adjutant General's Office. "Personnel Research Section, AGO: Development and Current Status." *Psychological Bulletin*, XLII (1945), 445-452.
5. Staff, Personnel Research Section, AGO. "The Army General Classification Test." *Psychological Bulletin*, XLII (1945), 760-768.
6. Staff, Personnel Research Section, AGO. "The Army General Classification Test with Special Reference to the Construction and Standardization of Forms 1a and 1b." *Journal of Educational Psychology*. (In press.)
7. Stewart, N. "Relationship Between M.O.S., A.G.C.T. Standard Score and Other Variables." *Personnel Research Section Research Study No. 675*, Part III. September 26, 1945 (mimeographed).
8. Stewart, N. "A.G.C.T. Scores of Army Personnel Grouped by Occupation." *Occupations*, XXVI (1947), 5-41.



# THE VALIDITY OF A MULTIPLE APTITUDE TEST AT THE SECONDARY SCHOOL LEVEL<sup>1</sup>

DAVID SEGEL  
U. S. Office of Education

THIS is a study of a multiple-aptitude test battery designed to measure relatively independent mental traits. In this battery are tested Word Fluency, Language Fluency, Reasoning, Spatial Relationships and various aspects of so-called "Mechanical Aptitudes." The multiple-aptitude test reported upon here was a battery adapted from aptitude tests in the War Department.<sup>2</sup> The data presented were obtained from the administration of the adapted tests to representative high-school populations.<sup>3</sup> The writer, in planning and carrying through this study, was motivated by the belief that one of the next important steps in the growth of measurement work at the secondary-school level would be the use of multiple-aptitude tests.<sup>4</sup> The fundamental reason for this belief lies in the possibilities for these tests to diagnose more fundamental strengths and weaknesses than other types of tests. For guidance purposes at the high-school level there have been three types of tests measuring the intellectual functions. Achievement tests—for specific subjects and in test batteries—are one of these types. These tests are valuable since they measure the actual accomplishment of what is taught in school. Intelligence tests are another of these types of tests. Intelligence tests usually

---

<sup>1</sup> This study was carried on in the Secondary Division of the U. S. Office of Education.

<sup>2</sup> Permission to adapt and to try out these tests is contained in an agreement between the Office of the Secretary of War and the Office of Education.

<sup>3</sup> The tests were given in Winston-Salem, N. C., Atlanta, Ga., Ardmore, Pa., and Philadelphia, Pa.

<sup>4</sup> Since beginning this study Traxler (*EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 3-16) has indicated the same belief expressed by the writer.

give only one score—a score on general mental ability. At the secondary level the value of general intelligence tests—without subtests giving scores on different traits—will become limited as multiple-aptitude tests are better developed. The third type of test developed for pupil personnel work is the prognostic test for a specific high-school subject such as for Algebra, a foreign language, English, Physics, etc. Such specific prognostic tests are the most valid for predicting success in the specific subject involved. However, due to the large number of subjects in the secondary school and the need for prognostic tests which predict success in more general areas such as Social Studies, Mathematics and Science, Languages, etc., the growth in the use of specific subject prognostic tests is limited. As multiple-aptitude tests are developed such specific tests will be used mainly for the sectioning of classes for instructional purposes. The multiple-aptitude test—involving as it does a measure of several mental traits—should be of value because it breaks down the general mental ability score so that the type of traits the student has is made known. Scores on different traits should be differentially related to success in different subjects. Such scores should be, therefore, of greater value than the all-round scores on a general mental ability test. Such multiple-aptitude tests should be of more value than achievement tests in predicting success in school and occupations because the traits tested are more fundamental than the tested achievement of school work.

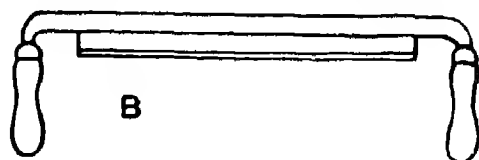
The plan of the investigation involved the administration of the aptitude tests to representative school populations in order to determine the adaptability of the tests to school populations through determining mean scores, standard deviations, and reliability coefficients and through an investigation of the validity of the tests for use in secondary-school guidance work. The population for any one table in this study is always based on the population in one school system. The population for different tables is usually for different populations. It is believed that the conclusions drawn are not affected by this limitation of the data.

The tests selected and adapted for use were as follows:

*Test No. 1. Mechanical Aptitude Tests*

Part 1. *Tool Usage*.—A sample question from this test is:

8.

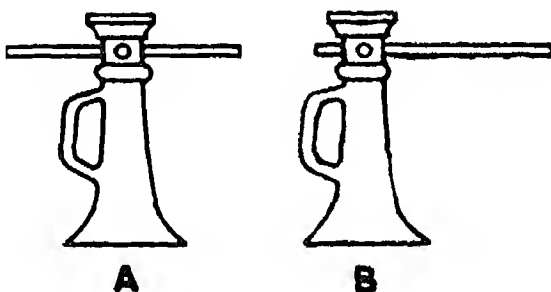


is used by a

- A) butcher
- B) carpenter
- C) draftsman
- D) machinist

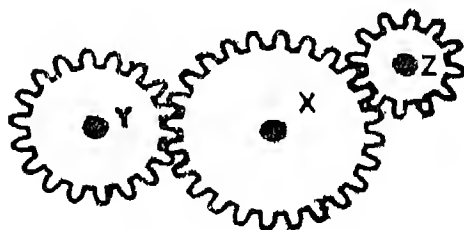
Part 2. *Mechanical Problems*.—A sample question from this test is:

54.



Which jackscrew will have greater lift with equal effort?

Part 3. *Pulleys and Gears*.—A sample question from this test is:

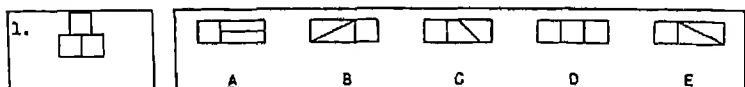


- 90. When the gears are moving, does Y turn faster than Z?
- 91. When the gears are moving, does Z turn in the same direction as Y?

*Test No. 2. Spatial Relationships Tests*

Part 1. *Two dimensional space relationships.*—In this test the directions and a sample question were:

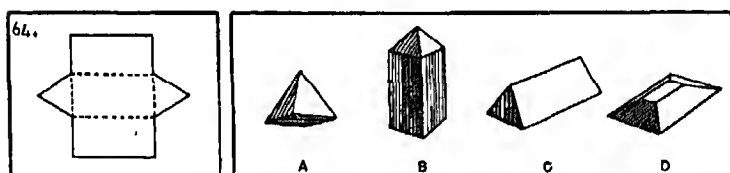
In this test the questions have numbered figures on the left and five lettered figures on the right. You are to find the one lettered figure which is made of exactly the same pieces as the numbered figure. Some of the pieces are turned around. Some are turned over. Here are some sample problems to show you how to make your answers on the answer sheet.



In problem 1 the number figure is made of three small squares. Only figure D has exactly the same three pieces. Now look at your answer sheet. After number 1, there are five squares labeled A, B, C, D, and E. A check has been made in the square labeled D since the D answer is the right one for question 1

Part 2. *Three dimensional space relationships.*—In this test the directions and a sample question were:

In this test you are to find the figure the pattern will make. Here are some sample problems.



In problem 64 if the flat piece of metal at the left were folded upward on the dotted lines it would like figure C. Now look at your answer sheet. After the number 64 there are four squares labeled A, B, C, and D. Check C since the C answer is the right one for problem 64.

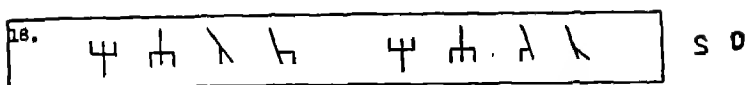
*Test No. 3. Speed of Perception*

Part 1. *Alphabetical Relationships.*—This is a test of classifying names of persons alphabetically.

Part 2. *Name and Number Comparison.*—This is a test comparing names and numbers to see if they are the same or different.

Part 3. *Visual Perception.*—This is a test to discover whether certain patterns are similar or dissimilar. A sample from this test is:





In this test the student is asked to compare the pattern of the first four figures with the last four and decide whether they are the same or different.

#### Test No. 4. Code Learning

A description of this test is made easiest by giving the directions which were:

This is a test of your ability to learn a CODE. In the practice exercises below is a CODE consisting of five symbols, each of which stands for a given number. You are to place in the box beneath each symbol the number which that particular symbol represents. The first box is correctly filled in. Do the rest of the exercise now.

CODE

○	□	△	=	0
4	2	1	5	3

EXERCISE

□	=	○	0	○	=	△	0	□	○	△	=	0	□	△
2														

The correct answers for the practice exercise are 5, 4, 3, 4, 5, 1, 3, 2, 4, 1, 5, 3, 2, 1.

#### Test No. 5. Word Fluency

This is a multiple-choice vocabulary test.

#### Test No. 6. Language Usage

This is a multiple-choice test of applied grammar, sentence structure.

#### Test No. 7. Mathematical Reasoning

A sample from this test is:

22. A student has an average grade of 85 in four courses. What grade must he get in a fifth course so that his average for all five will be 87?

#### Reliability and General Adaptability of the Tests

Table 1 gives for each test the testing time, the number of items in the tests, the standard deviation and the mean score. Table 2 gives the reliability coefficients for certain of the tests. These are all based on 100 cases in the 10th grade. The mean

TABLE 1

*Data on Time, Number of Items, Mean and Variability of Scores on Tests  
(100 cases 9th Grade)*

	Time	Number of items	Mean score	Standard deviation
Test No. 1—Mechanical Aptitude Tests				
Part 1—Tool Usage .....	8 min.	46	27.2	8.10
Part 2—Mechanical Problems .....	8 min.	24	15.1	3.80
Part 3—Pulleys and Gears .....	8 min.	50	23.4	6.02
Test No. 2—Spatial Relationships Tests				
Part 1—Two dimensional space relationships .....	9 min.	50	26.7	10.78
Part 2—Three dimensional space relationships .....	9 min.	40	23.3	5.58
Test No. 3—Speed of Perception				
Part 1—Alphabetical relationships ..	5 min.	148	76.7	24.15
Part 2—Name and number comparison ..	7 min.	200	97.1	19.00
Part 3—Visual perception .....	4 min.	44	32.0	5.10
Test No. 4—Code Learning .....	8 min.	420	97.6	19.00
Test No. 5—Word Fluency .....	7 min.	50	27.4	6.92
Test No. 6—Language Usage .....	10 min.	36	14.3	5.48
Test No. 7—Mathematical Reasoning ..	17 min.	32	16.8	4.26

scores for juniors and seniors were slightly higher, but the differences were not great. These data show that most of the tests are adapted to high-school students. The reliability coefficients (Tests 1, 5, 6, and 7) are high for the number of items in the tests. The reliabilities for Tests 2, 3, and 4 are high, but they were not calculated since the test items are nearly identical and would result naturally in high reliability coefficients. With lengthier tests, especially for Tests No. 5, Word Fluency, No. 6, Language Usage, and No. 7, Mathematical Reasoning, these tests can be brought up to a reliability substantially above .90 and could therefore be used in individual diagnosis.

TABLE 2

*Reliability Coefficients for Tests 1, 5, 6, and 7*

Test 1	
Part 1 .....	.84
Part 2 .....	.58
Part 3 .....	.81
Test 5 .....	.83
Test 6 .....	.69
Test 7 .....	.58

*Validity of the Tests*

The validity of these tests has been investigated in three ways: (1) to show the power of these tests to predict success in various high-school subject areas; (2) to show that the tests

TABLE 3

*Original Correlation Coefficients\* for High-School Subject Areas and Results on Tests 1, 5, 6, and 7 in One School System (11th grade level)*

Subject Area	Test 1 Mechanical Aptitude Tests	Test 5 Word Fluency	Test 6 Language Usage	Test 7 Mathe- matical Reasoning
Industrial Arts ..... (87 cases)	.48	.09	.12	.26
Foreign Languages ..... (78 cases)	-.04	.20	.54	.32
Biology ..... (70 cases)	.07	.38	.54	.24
Social Studies ..... (112 cases)	.25	.30	.10	.23
English ..... (120 cases)	.02	.49	.32	.11
Mathematics ..... (104 cases)	.20	.19	.30	.62

\* Coefficients in italics are less than three times their probable error.

TABLE 4

*Correlation Coefficients between High-School Subject Areas and Results on Tests 1, 5, 6, and 7 Corrected for Attenuation in the Criterion*

Subject Area	Test 1 Mechanical Aptitude Tests	Test 5 Word Fluency	Test 6 Language Usage	Test 7 Mathe- matical Reasoning
Industrial Arts ..... (87 cases)	.68	.13	.17	.37
Foreign Languages ..... (78 cases)	.00*	.28	.76	.45
Biology ..... (70 cases)	.10	.54	.76	.34
Social Studies ..... (112 cases)	.35	.42	.14	.33
English ..... (120 cases)	.03	.70	.45	.16
Mathematics ..... (104 cases)	.28	.27	.42	.88

\* This correlation coefficient was negative in the original table. If positive relationships are assumed between all measures it seems reasonable to assign zero as its value.

are relatively independent and can therefore be used in a differential diagnosis; and (3) to compare the relative efficiency of these tests with other tests which appear to be measuring similar traits.

(1) For the first classification of the evidence for validity, Tables 3, 4, 5, and 6 are presented.

TABLE 5

*Correlation Coefficients\* between Marks in Certain High-School Subjects and Results on Tests 2, 3, 4, 5, 6, and 7 (57 cases, 9th grade level)*

	English	Mathematics and Science	Commercial
Test No. 2—Spatial Relationships Tests			
Part 1—Two dimensional space relationships .....	.07	.35	.03
Part 2—Three dimensional space relationships .....	.09	.10	-.02
Test No. 3—Speed of Perception			
Part 1—Alphabetical relationships .....	.37	.30	.13
Part 2—Name and number comparison .	.21	.10	.20
Part 3—Visual perception .....	-.03	-.13	-.04
Test No. 4—Code Learning .....	.13	.38	-.02
Test No. 5—Word Fluency .....	.41	.19	-.09
Test No. 6—Language Usage .....	.50	.21	.20
Test No. 7—Mathematical Reasoning .....	.36	.42	.18

\* Coefficients in italics are less than three times their probable error.

TABLE 6

*Correlation Coefficients between High-School Subject Areas and Results on Tests 2, 3, 4, 5, 6, and 7 Corrected for Attenuation in the Criterion*

	English	Mathematics and Science	Commercial
Test No. 2—Spatial Relationships Tests			
Part 1—Two dimensional space relationships .....	.10	.50	.04
Part 2—Three dimensional space relationships .....	.13	.14	.00*
Test No. 3—Speed of Perception			
Part 1—Alphabetical relationships .....	.52	.42	.18
Part 2—Name and number comparison ..	.30	.14	.30
Part 3—Visual perception .....	.00*	.00*	.00*
Test No. 4—Code Learning .....	.18	.54	.00*
Test No. 5—Word Fluency .....	.58	.27	.00*
Test No. 6—Language Usage .....	.71	.30	.30
Test No. 7—Mathematical Reasoning .....	.51	.60	.25

\* These correlation coefficients were negative in the original table. If positive relationships are assumed between all measures it seems reasonable to assign zero as their values.



To best understand the results, examine Tables 4 and 6. These are the tables of original coefficients corrected for attenuation. These tables show more accurately the relationships of these test results to the subject areas than do the original tables.<sup>5</sup> These tables show substantial predictive power for the different subject areas for such short tests.

(2) Next is presented the table of intercorrelations between the tests (Table 7). If these intercorrelations are compared with the reliabilities of the individual tests (Table 2) and with Tables 4 and 6 it will be seen that the coefficients in general indicate significant differences between the traits measured.

(3) Other studies have not been made of the predictive values of such multiple-aptitude tests at the high-school level. The nearest comparison to be made is with those on the college level. Studies on the college level which may be cited are those by Harrell and Faubian<sup>6</sup> and by Bernreuter and Goodman.<sup>7</sup> The latter study is on the college freshman level and the best for comparison with the results for high-school students presented in this paper. Table 8 gives the results of the Bernreuter and Goodman study.

TABLE 8

*Table Showing Correlation between the Results of the Thurstone Primary Abilities Tests and Marks in College Courses (After Bernreuter and Goodman)*

Primary Abilities Test	Semester Average	Chemistry	English Composition	Mathematics
P	.04	.07	.05	.04
N	.32	.27	.26	.27
V	.33	.32	.44	.16
S	.23	.19	.11	.25
M	.10	.04	.23	.05
I	.34	.23	.21	.29
D	.38	.41	.21	.44

<sup>5</sup> For those interested in the reason for this statement, see Segel, David. "Validity of the V. C. Aeronautics Aptitude Test and the O. E. Scientific Aptitude Test." *Journal of Psychology*, XVIII (1944), 65-80.

<sup>6</sup> Harrell, Willard, and Faubian, Richard. "Primary Mental Abilities and Aviation Maintenance Courses." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, I (1941), 59-66.

<sup>7</sup> Bernreuter, R. G. and Goodman, C. H. "A Study of the Thurstone Primary Abilities Test Applied to Freshmen Engineering Students." *Journal of Educational Psychology*, XXXII (1941), 55-60.

Compare the results of Table 8 with those of Tables 3 and 5. In making this comparison keep in mind that the Thurstone Primary Abilities are much longer tests than those in the Office of Education battery. The conclusion, then, is that the battery of tests used by the Office of Education is as efficient if not more efficient than are other aptitude tests for guidance purposes.

### *Conclusions*

1. The type of material in certain War Department aptitude tests can be used with high-school students in grades 9 through 12.
  2. If tests of the type presented were lengthened, their reliability could easily be brought above .90, or any needed satisfactory level.
  3. The evidence in this study supports the hypothesis that a multiple-aptitude test of this type is of value for differential diagnosis and prognostic work.
-





## STUDENT PERSONNEL WORK IN NATIONAL SETTING

C. GILBERT WRENN

*University of Minnesota*

THE attention devoted to the student personnel function at the National Conference on Higher Education, Chicago, March 31-April 3, 1947, demonstrates the current emphasis on this phase of higher education. Of the nineteen groups into which the approximately five hundred delegates to this conference were divided, two, Group Ten on "Counseling" and Group Eleven on "Student Welfare," considered the total program of personnel services on the campus, while two others considered the more specific student personnel topics of Admissions and Veterans Advising.

The resolutions which form the body of this article were submitted by Group Ten, of which the writer was Chairman. This group was composed of about fifty delegates from almost that same number of institutions. The group worked intensively for two days, not upon the resolutions but upon a rather basic discussion of trends in college personnel work. Several sub-committees brought in reports on specific phases of the total area; the in-service training of counselors, the counseling of veterans, the administration of student personnel work, and the evaluation of personnel services. (The names of the chairman of these committees and other representatives of the group are listed at the end of this article.) The total proceedings of this group effort, as is true of each of the other eighteen groups, have been released in a volume published by the Department of Higher Education of the National Education Association, the organizer and sponsor of the Conference. Each group has one chapter in the volume, written by the Recorder for his group.

The resolutions give attention to what were considered current crucial problems in the total field. They are in such discrete areas as the coordination of occupational information and the coordination of measurement programs on a national scale, standards for the training of student personnel workers, the conservation of veteran's guidance centers, and a petition to the Congress on their action regarding educational issues. This discreteness is more apparent than real if it is realized that consideration had been given during these two days to the complete scope of student personnel work.

- The report submitted to the total Conference by its Chairman on the final day of the Conference contained the following conclusions on the broad program in this area of higher education.

I. The basic assumptions upon which a student personnel program operates are as follows:

1. The student is a unified organism and the development of his total personality is the concern of the institution.
2. Personnel services and instructional services both contribute to the rounded development of this personality and their combined functions form the educational program of the institution.
3. The philosophical point of view under which both instructional and personnel services can function with the optimum benefit to the individual student, considers each student as unique, is concerned with the development of the total personality, and develops the educational program of the student in terms of his present interests and needs. This point of view is dedicated to meeting the needs of an individual student as these exist within the framework of the demands and needs of contemporary society.

II. Provision must be made for the coordinated administration of all personnel functions under an administrative head responsible directly to the president upon a basis parallel to that of the administrative head of the instructional functions and the head of the business functions of the institution. The personnel administrator has a direct 'line' responsibility for some personnel functions, and a staff relationship to others. . . .

III. Counseling is one of the several functions of a personnel program. Other functions include those of orientation of new students, supervision of living arrangements, provision and supervision of student activities, health services, part-time employment and placement services, and maintenance of

personnel records. Counseling may be a phase of many of these services, but counseling designed to aid the student in his immediate adjustment and to give him skilled assistance in planning for the future is a major function in itself. Such counseling calls for skilled appraisal of the individual's aptitudes and interests and for a thorough understanding of his emotional conflicts and development. It is a function engaged in by faculty on a casual or on a part-time basis and by counselors whose training and skill are adequate for full-time application. Counseling is performed at different levels of intensity by these various educational workers, from the counseling which is a function of good instruction to that which deals with complex problems of adjustment and life planning. . . .

IV. The regulation of student conduct, in harmony with modern concepts of mental hygiene, calls for diagnosis and remedial action upon a clinical counseling basis, with any disciplinary action to be taken only after remedial action has failed. . . .

V. The evaluation of student personnel procedures is an integral part of the program. New procedures should be established in such a manner that their evaluation is possible and decision as to continuance or modification should be upon the basis of proved value. . . .

## RESOLUTIONS

BE IT RESOLVED, that immediate steps be taken to implement the following recommendations on problems which require cooperative study and action among colleges and universities and national agencies:

1 *Occupational Information in Fields of Work Employing College and University Graduates.* It is recommended that a cooperative study be made of the most effective means:

a) of collecting and disseminating to colleges and universities usable information concerning occupational conditions, opportunities and trends in those fields of work employing college and university graduates, and

b) of applying this information to the problem of relating the supply of trained personnel to the demands for their services in American society.

It is suggested that the study be planned and administered by a joint Committee composed of designated representatives of the NEA, ACE, NVGA, U. S. Office of Education, Occupational Outlook Service, and of representative college authorities, the Committee to report its findings to the associated agencies and later to colleges and universities generally.

2. *Measurement Programs.* It is recommended that co-ordinative means be sought to pool the resources and information of national testing agencies, governmental and non-governmental, in the following areas:

a) Provision of improved means of appraising individuals and of giving vocational orientation to candidates for advanced instruction in those fields of training commonly represented in American colleges and universities.

b) Extraction from war service programs of those measurement devices and procedures most suitable for civilian use and their application to the clinical problems of counseling veterans and students

c) Study of the most suitable means of selection and guidance of youth receiving federal subsidies in the field of higher education. It is suggested that, in this connection, membership of this department study the proposals, recently made, to coordinate the services of certain test agencies serving the college field.

3. *Veterans Guidance Centers.* It is recommended that a joint study be immediately inaugurated by national educational groups and associations, of the social and professional values served by the nationwide system of veterans guidance centers now operative in more than 350 colleges and universities in this country; that attention be given to the need for conserving such present values and facilities of veterans guidance centers as: experienced and professional personnel who are acquainted with local needs and conditions, established referral channels within the community, and the accumulated stock of diagnostic materials with local norms; that careful exploration be made of the continuing need for guidance centers under educational auspices for assisting individuals and groups with their adjustment and orientation problems; and that the feasibility be examined of financing on a federal-state-local basis, without federal control, of a national system of guidance centers under educational auspices to serve individual needs in communities and regions.

4. *Qualifications of Student Personnel Workers.* It is recommended that the Committee on Student Personnel Work of the American Council on Education inform the executive officers of the national educational groups (and their constituent memberships) of salient developments in the preparation and certification of student personnel workers such as a *Manual of Counselor Training* now being prepared by the Division on Professional Training and Certification of the National Vocational Guidance Association; that colleges and universities engaged in professional training of counselors secure, through the American Council on Education Committee, the Division of Personnel and Guidance of the American Psychological Association, and the Council of Guidance and Personnel Associations, the most effective collaboration of members, and that means be adapted through the respective educational associations of informing all administrators of the development of the proposed *Manual* and of other applications to the problems of personnel selection.

5. *Joint National Advisory Action on Education Legislative Issues.* It is recommended that with respect to action by the Congress and its committees on existing activities of Federal Government affecting educational interests that:

a) All appropriations affecting educational welfare should be considered in the light of the total interrelated needs and demands, and that action such as that recently taken by the House Appropriations Committee which wiped out such services as the National Clearing House and the Outlook Service of the Department of Labor not be taken without careful consideration of the permanent educational values lost by such action;

b) The needs for equalized educational opportunity throughout the country and of raising the standards of living of all education staff workers are of paramount importance;

c) Wise national economies should be planned and practiced consistent with maintenance of essential services;

d) Decisions as to the worth of specific services within the total complex of need may best be made through the joint judgments of the appointed leaders of representative national agencies in the field of education;

e) It is recommended that this Conference go on record favoring that every means be utilized to inform state legislatures and college boards of trustees of the nature, meaning and importance of student personnel work.

*Group Ten on Counseling and Other Student  
Personnel Services*

Chairman—C. Gilbert Wrenn (Minnesota)

Recorder—Mrs. Herbert Hawkes (Mills College)

Steering Committee: W. W. Blaesser (Washington State)

Wray H. Congdon (Lehigh)

George Hilliard (Western Michigan)

Henry B. McDaniel (Stanford)

Robert H. Mathewson (Harvard)

Chairmen of subcommittees: Administration, Henry B. McDaniel (Stanford); Evaluation, John R. Beery (Miami); In-Service Training, John H. Cornehlson (San Francisco State); Veterans Counseling, George D. Small (Tulsa); Resolutions, Robert H. Mathewson (Harvard).



# MEASURING OUTCOMES OF CLASSROOM THERAPY<sup>1</sup>

HAROLD B. PEPINSKY<sup>2</sup>

Michigan State College

THE purpose of this paper is to discuss (1) the feasibility of individual therapy through classroom groups and (2) experimentation in progress at Michigan State College.

## *Historical Background*

The experimental study of individual therapy through classroom groups<sup>3</sup> has its origin in four major approaches to the study of human relationships. A first major approach is *group therapy with deviant individuals*; a second, *group therapy with normal individuals*; a third, *student personnel work in educational institutions*; and a fourth, *the study of interpersonal relationships in social groups*.

## *Group Therapy with Deviant Individuals*

In the past five years psychiatrists and clinical psychologists have turned with increasing frequency to the social group as a medium for treating psychosomatic disorders among severely maladjusted patients. The exigencies of a world war brought into sharp focus the impossibility of treating large numbers of patients by individual methods alone; the pioneer work of Jacob Moreno and his associates had indicated rich possibilities in group methods (14). As a reflection of growing interest in

---

<sup>1</sup> From a paper, "A Classroom Experiment in Group Therapy," addressed to members of the American College Personnel Association, Columbus, Ohio, March 30, 1947.

<sup>2</sup> Grateful acknowledgment is due Dr. Leo A. Haak, Head of the Basic College Department of Effective Living, for making possible the present experimental program and for numerous helpful suggestions.

<sup>3</sup> I am indebted to Dr. C. E. Erickson for suggesting that "individual therapy through groups" reflects more appropriately than "group therapy" a concern with the adjustment of the individual

group psychotherapy, one may note the number of articles under the heading "Group Psychotherapy" listed in *Psychological Abstracts* for the years 1942-1946.

TABLE I  
*Articles on "Group Psychotherapy" Listed in Psychological Abstracts  
for the Years 1942-1946*

<i>Year</i>	<i>Number of Articles</i>
1942	3
1943	6
1944	9
1945	23
1946	40

The number of articles so listed has steadily increased from three in 1942 to forty in 1946.

### *Group Therapy with Normal Individuals*

For many years the social group worker has been actively engaged in preventive and remedial therapy especially with underprivileged groups. Today under the leadership of Fritz Redl (23) and S. R. Slavson (28) group workers are making numerous applications to normal group situations of clinical practices derived from the study of deviant groups.

### *Student Personnel Work in Educational Institutions*

Two major developments in colleges and secondary schools should be noted in connection with the historical background of the present experiment. One is the emergence of *individual clinical counseling* as a professional field, and the other is the growth of the more generalized field of *guidance*. From clinical counseling has come an important synthesis of clinical and experimental methods in the diagnosis and treatment of minor functional maladjustments (18). From the generalized field of guidance have come important supplemental group guidance techniques (1, 10, 33). I have used the term "supplemental" advisedly in speaking of group guidance techniques since this is an area in which much practical work has been done (7, 16) as against very little experimentation.<sup>4</sup>

<sup>4</sup> Caution in accepting present results and need for further experimentation is emphasized by the studies of Stone (31) and Sachs (25)



*The Study of Interpersonal Relationships  
in Social Groups*

No study of individual human behavior in classroom groups can be adequate without reference to the stimulating and significant contemporary research in interpersonal and intergroup relationships. A basic pattern for research in this area is the sociometric and sociographic approach of Jacob Moreno (15) with further refinement of techniques provided by Helen Hall Jennings (11). Recent applications of sociometric and sociographic method to the analysis of interpersonal relationships in schoolroom and extra-curricular life (6, 17, 19, 20, 21, 22, 30) appear to have extremely important implications for the student personnel worker. I shall return to this in my discussion of experimental methodology.

A second major contribution to the study of interpersonal relationships has been made by Harold Anderson and his associates (3, 4) in their studies of "dominative and socially integrative behavior" of classroom teachers. The measured effects of teachers' classroom personalities upon the behavior of children in the classroom have profound implications for the student personnel worker. New and stimulating classroom techniques, documenting an awareness of these implications, have been suggested by Cantor (5), Pressey (20), Pressey and Hanna (21), and Shedlin (27).

A final major contribution has been made in the field of "action research" in which prejudiced attitudes and behavior are studied in actual social situations (12). Reading the excellent brochure "Improving Inter-group Relations in School and Community Life" (9) published last year by the North Central Association, one is again impressed by the extent to which an individual's adjustment is affected by his status in the primary social group, and by the extent to which this adjustment in turn is affected by the status of his basic group in school and community life.

The present study of individual therapy through classroom groups has drawn freely from the above-mentioned major sources in inspiration and in technique.

*The Experimental Setting*

The project on individual therapy through classroom groups is being conducted in the first quarter of a three-quarter course in "Effective Living" at Michigan State College. A brief description of the course and of its place in the all-college program will give a clearer picture of the experimental setting.

Effective Living is one of the seven "basics" comprising the Basic College Program at Michigan State College. Each "basic" is a three-quarter general education course, and all students graduating from Michigan State College are required to have completed five of the seven "basic" courses in addition to the specialized courses in their major fields.

The first quarter of the course in Effective Living deals with basic values in human relationships, the second quarter with the individual's adjustment to marriage and family living, and the third quarter with the individual in his larger group relationships.

In the first quarter the student attends a general lecture section for one hour each week and an hour discussion-laboratory section three times a week. Lecture sections are comprised of approximately 300 students, with a maximum of 35 students in each discussion-laboratory section. It is in certain of the smaller sections that experimentation in classroom therapy is taking place.

The purposes of the first quarter's work are to help the individual student: to define basic values toward which men strive; to evaluate these in terms of our time and cultural setting; to take stock of his own behavior and values; and to acquire practical and systematic knowledge necessary for making progress toward attaining culturally and personally desirable goals and values. Although the course in Effective Living cuts across traditional subject matter fields, considerable evidence is brought to bear upon the above problems from the major fields of Contemporary Philosophy, Psychology and Sociology.

The Department of Effective Living staff is laboring under no delusions in appraising the magnitude of the task which it has set for itself. Experimental evidence from studies made in similar courses concerning attitude changes effected through

the medium of the classroom is far from encouraging whether such evidence is drawn from subject-centered or student-centered courses (8, 26). Recent enthusiastic reports on the application of nondirective teaching techniques (5, 27) are encouraging yet must be accepted with caution pending further investigation.

At Michigan State College, however, as at other institutions of higher education throughout the country, postwar enrollment increases have been alarmingly great. From a pre-war peak of 7,000 students, the enrollment jumped to more than 15,000 in the Fall Quarter, 1947, and shows no signs of falling off for several years. Although the individual counseling staff of the College now includes more than a dozen full-time workers, this number cannot reach more than a small fraction of the students needing counseling assistance. Therefore, the Effective Living Department, because of the nature of its course materials and the pressure of its great number of students, has been forced to assume supplemental counseling functions apart from any desire on its part to do so.

In the realization that no single agency on a large college campus can hope to do the job of student personnel work alone, departmental policy has been to promote friendly working relationships and confidential exchange of information about students with other academic departments and service agencies on the campus and with numerous interested outside groups. Particular attention has been given to the coordination of departmental services with those of the Counseling Center. Last year a member of the Counseling staff taught a discussion section in the first quarter of the Effective Living course as a further means of promoting harmonious working relationships.

In addition, active interchange of ideas and services is maintained with the Institute of Counseling, Testing and Guidance. This agency has been established to coordinate training, research, and off-campus consultant activities in personnel work at Michigan State College. During the Spring Quarter, 1947, and through the Institute, a graduate student majoring in Student Personnel Work was assigned to the Effective Living Department for supervised field work in the study of individual

therapy through groups. Similar assignments are in prospect for 1947-48. Moreover, several staff members are working toward doctorates in personnel work.

A second departmental policy has been the promotion of harmonious working relationships among the members of the departmental staff. This policy invites the active participation of each staff member in departmental activities and planning and is based on the hypothesis that a sense of individual belongingness is essential to wholesome departmental growth (2). Such a system of "integrative" relationships tends to have as one of its end results a better classroom atmosphere (4).

A third departmental policy calls for an active evaluation program as a means of predicting and controlling the outcome of classroom activities. An important question is the extent to which students adjust more favorably to life situations as a result of having taken the course. The experimental study of individual therapy through groups is one attempt to answer this question.

#### *Assumptions Underlying the Experimental Program*

The extent to which students adjust more favorably as a result of individual therapy through classroom groups is not a simple determination but involves the testing of numerous hypotheses. Before proceeding to the hypotheses, however, let us examine some of the assumptions which underly the present experimental program:

1. The goals of individual therapy and of individual therapy through groups are the same, i.e., to assist the individual in solving his adjustment problems.

2. Since there is no definitive evidence to support the exclusive use of one or another mode of therapy, we are justified in continuing to experiment (13). The present study is justifiable on this basis.

3. With limited facilities for individual counseling, the group approach offers larger numbers of students access to the therapist.

4. The classroom group is a projective medium through which the individual student may express his own ideas and

actions, with the possibility of seeing these reflected in the attitudes of group members toward him.

5. The classroom groups in which individual therapy is to be attempted are organized on an informal discussion basis. The content of the course is relevant to a discussion of human behavior. The first quarter of the course in Effective Living has been designed to satisfy this assumption.

### *Hypotheses to be Tested in the Experimental Program*

The list of hypotheses proposed below is far from complete. It is anticipated that others will be suggested as experimentation progresses. The central hypothesis is, of course, that students can be aided to make better life-adjustments as the result of individual therapy through classroom groups.

1. Through the medium of therapeutic classroom sessions individual students can be motivated to seek help in solving their personal adjustment problems. Many clinicians are agreed with Rogers (24) that desire for help by the client is a prerequisite to effective treatment.

2. Individual students can gain self-insight by testing their behavior against the reactions of the classroom group. A subsidiary hypothesis is that transfer of learning in this situation can take place to new group situations.

3. Individual students can gain reassurance in accepting their adjustment problems through identification with problems revealed by others in the group.

4. Accuracy of individual diagnosis by the therapist can be increased through the additional study of individual students in the classroom group.

5. Individual students can acquire needed information (e.g., information relevant to making appropriate educational-vocational choices) through the medium of the classroom group.<sup>5</sup>

6. Individual students with certain kinds of adjustment

---

<sup>5</sup> It should be noted that in Stone's (31) exploratory study, informal discussion groups did not necessarily accompany the large lecture sections. This is particularly significant in view of Stone's finding that students taking a vocational orientation course alone were apt to make inappropriate choices. Moreover, we do not know to what extent Stone's findings are a function of the classroom procedure employed

problems, e.g., lack of social skills, can be aided to make better adjustments through the classroom group than can students who have had individual counseling alone (20, 21).

7. Larger numbers of students can be aided through combined individual counseling and classroom therapy than through individual contacts alone.

8. Effective individual therapy through classroom groups requires special clinical skills and group leadership ability in the therapist.

### *Experimental Methodology*

In order to test hypotheses relevant to the present experiment, a sound experimental methodology must be developed. It should be emphasized that no definitive answers can be given without reference to findings based on *experimental* and *control* groups. Such groups should consist of a control group of students who have had no formal therapy and experimental groups of students who have had individual counseling, classroom therapy, or both. One needs further to consider differences resulting from variation among therapists, from variation in kinds and severity of maladjustment, etc. These and other variables make an inclusive experimental design well-nigh impossible at the present time.

But the difficulty of testing an inclusive hypothesis does not permit one to sidestep investigation in this area. Although one admits the complexity of the larger problem, it is still possible to make a beginning: (1) by measuring behavior changes in students who have taken the first quarter of the course in Effective Living; (2) if such changes occur, by determining the direction of behavior change; and (3) by testing to see whether such behavior changes differ measurably from those of students who have not taken the course in Effective Living.

Two pertinent sources of information concerning behavior change among students in the classroom are: (1) the way in which an individual pictures himself (*psychological* make-up), and (2) the way others picture him (*psychosocial* make-up). The self-picture of an individual student may be obtained partially through use of such instruments as the *Kuder Preference Record*, the *Allport-Vernon Study of Values*, and the *Minne-*

*sota Multiphasic Personality Scale.* The picture which others have of the individual student may be obtained by means of sociometric and sociographic methods to determine whether he is accepted, rejected, or simply not noticed in the group. Of course, it may be argued that an individual's picture of himself is inseparable from the way he is pictured by others, yet methodologically we must distinguish between what he sees and what is seen by others.

Using these two kinds of measures, we can obtain pictures of an individual student before and after he has taken the first quarter of the course in Effective Living. Differences between these pictures may be interpreted as changes in verbal behavior by and toward the student. This method may be supplemented by individual case studies of deviant members of the classroom group. Such case data are already being collected and bring to light many fruitful hypotheses.

At the present time a preliminary investigation is being made of the relationship between psychological and psychosocial measures to determine whether an individual's picture of himself is congruent with the picture others have of him. Specifically, we are interested in an individual's *Minnesota Personality Scale* scores as against sociometric ratings by other students.

From this level of descriptive study of students in the classroom it will be possible to evaluate changes in students after they have taken the course. The next step will be to collect data on students in different discussion sections, students with different instructors, and students who have not taken the course in Effective Living. These data will provide a basis for a long term evaluation of the outcomes of individual therapy through classroom groups. With a group of competent judges and with adequate case data on individual students, it would then be possible to assess the total adjustment of experimental and control groups of students at the end of one, five or ten years.<sup>6</sup>

### *Summary and Conclusions*

Because of limitations in a short introductory paper cover-

---

<sup>6</sup> This use of a judgment criterion has been recommended by Williamson and Bordin (32).

ing a wide area of possible experimentation, the statement of basic assumptions and tentative hypotheses has been presented without the kind of elaboration that is really needed. What has been attempted here is a discussion of the feasibility of individual therapy through classroom groups with particular reference to experimentation in progress at Michigan State College. Because the work is still in the planning stage, this presentation has been made in the hope that criticisms and suggestions will be forthcoming and that other investigators will be stimulated to begin work on similar problems. Research in this field seems to offer real possibilities for an increased effectiveness of therapy.

## REFERENCES

1. Allen, R. D. *Common Problems in Group Guidance*. New York: Inor, 1933.
2. Allport, G. W. "The Psychology of Participation." *Psychological Review*, LII (1945), 117-132.
3. Anderson, H. H. and Brewer, H. M. "Studies of Teachers' Classroom Personalities: I. Dominative and Socially Integrative Behavior of Kindergarten Teachers." *Applied Psychology Monographs*, No. 6, 1945.
4. Anderson, H. H. and Brewer, J. E. "Studies of Teachers' Classroom Personalities: II. Effects of Teachers' Dominative and Integrative Contacts on Children's Classroom Behavior." *Applied Psychology Monographs*, No. 8, 1946.
5. Cantor, N. *Dynamics of Learning*. Buffalo: Foster and Stewart, 1946.
6. Cook, L. A. "An Experimental Sociographic Study of a Stratified 10th Grade Class." *American Sociological Review*, X (1945), 250-261.
7. Dixon, F. B. "Group Guidance Techniques." *A Basic Text for Guidance Workers* (C. E. Erickson, Ed.). New York: Prentice-Hall, 1947. chap. 10.
8. Fattu, N. "Evaluation Program in Individual Orientation." *Studies in Higher Education*. Biennial Report of the Committee on Educational Research, 1938-1940. Minneapolis: University of Minnesota, 1941. pp. 108-130.
9. Harnly, P. W. *et al.* "Improving Intergroup Relation in School and Community Life." A Study conducted and reported by the Sub-Committee on the In-Service Education of Teachers, North Central Association of Colleges and Secondary Schools, 1946.
10. Hoppock, R. *Group Guidance Principles and Techniques*. (First manuscript, incomplete.) New York: Robert Hoppock, 1946.



11. Jennings, H. H. *Leadership and Isolation, a Study of Personality in Interpersonal Relations*. New York: Longmans, Green and Co., 1943.
12. Lippitt, R. and Radke, M. "New Trends in the Investigation of Prejudice." *Annals of the American Academy of Political and Social Science*, CCXLIV (1946), 167-176.
13. Mcchl, P. E. and McClosky, H. "Ethical and Political Aspects of Applied Psychology." *Journal of Abnormal and Social Psychology*, XLII (1947), 91-98.
14. Meiers, J. L. "Origins and Development of Group Psychotherapy; a Historical Survey, 1930-1945." *Psychodrama Monographs*, No. 17, 1946.
15. Moreno, J. L. *Who Shall Survive? A New Approach to the Problems of Human Interrelations*. Washington: Nervous and Mental Disease Publishing Co., 1934. (Now available through Beacon House, New York 17.)
16. Munson, J. B. "Organized Group Guidance Activities." *A Basic Text for Guidance Workers*, (C. E. Erickson, Ed.). New York: Prentice-Hall, 1947. chap. 11.
17. Olson, W. C. "The Improvement of Human Relations in the Classroom." *Childhood Education*, XXII (1946), 317-325.
18. Pepinsky, H. B. "Therapeutic Counseling." *A Basic Text for Guidance Workers*, (C. E. Erickson, Ed.). New York: Prentice-Hall, 1947. chap. 6.
19. Peters, M. L. "A Study of Social Acceptance, Rejection and Isolation among Children from the Second through the Tenth Grade." Unpublished doctoral dissertation, Ohio State University, 1945.
20. Pressey, S. L. "Social and Useful College Classes." *School and Society*, LV (1942), 117-120.
21. Pressey, S. L. and Hanna, D. C. "The Class as a Psycho-Sociological Unit." *Journal of Psychology*, XVI (1943), 13-19.
22. Rath, L. and Schweikart, E. F. "Social Acceptance Within Interracial School Groups." *Educational Research Bulletin*, XXV (1946), 85-90.
23. Redl, Fritz. *Recent Publications and Materials (1940-1944)*. Detroit: Wayne University (no date).
24. Rogers, C. R. *Counseling and Psychotherapy*. Boston: Houghton-Mifflin Co., 1942.
25. Sachs, G. M. *Evaluation of Group Guidance Work in Secondary Schools*. Los Angeles: University of Southern California Press, 1945.
26. Shaw, M. "Learning Study in Orientation." *Studies in Higher Education*. Biennial Report of the Committee on Educational Research, 1938-1940. Minneapolis: University of Minnesota, 1941. pp. 91-95.
27. Shedlin, A. J. "A Student-Centered Class." *Personnel Counselor*, II (1947), 116-131.

28. Slavson, S. R. *An Introduction to Group Therapy*. New York: Commonwealth, 1943.
29. Slavson, S. R. *Recreation and the Total Personality*. New York: Associated Press, 1946.
30. Smucker, O. C. "A Sociographic Study of the Friendship Pattern on a College Campus." Unpublished doctoral dissertation, Ohio State University, 1945.
31. Stone, C. H. "Evaluation Program in Vocational Orientation." *Studies in Higher Education*, Biennial Report of the Committee on Educational Research, 1938-1940. Minneapolis: University of Minnesota, 1941. pp. 131-145.
32. Williamson, E. G. and Bordin, E. S. "The Evaluation of Educational and Vocational Counseling: A Critique of the Methodology of Experiments." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, I (1941), 5-24.
33. Wrenn, C. G., Hein, R. J. and Pratt, S. *Aids for Group Guidance*. Minneapolis: Educational Publishers, 1942.

# THE FREQUENCY OF ERRORS IN THE CLASSIFICATION OF INDIVIDUALS ON THE BASIS OF FALLIBLE TEST SCORES

ROBERT L. EBEL

University of Iowa

## I. *Errors of Selection*

INDIVIDUALS are frequently selected for training or employment on the basis of test scores. Because these scores are not perfectly accurate, some individuals may be wrongly selected and others wrongly rejected. It is appropriate to ask how frequently such errors of selection are likely to occur in a given set of circumstances.

No direct answer to this question was found in the available literature on testing. Discussions of certain other aspects of the general problem of accuracy in selection are noted in the final section of this article.

The first step in estimating the frequency of errors of selection is to identify the factors which affect it. One of these is the reliability of the test scores used. With scores of low reliability, the frequency of errors would probably be relatively high. Another factor is the proportion of applicants selected. If only 5 per cent are selected, fewer errors would be expected than if 50 per cent of the applicants are selected.

In order to visualize the problem and to get some clues to its solution, let us assume that we have, for each of 100 individuals, both a fallible obtained score and an error-free true score on some classification test. These paired scores could be exhibited on a scattergram. If the distributions of both obtained and true scores are approximately normal, and if the coefficient of correlation between obtained and true scores is .92, the frequencies in the cells of the scattergram would be approximately those shown in Fig. I.

Now suppose that we are required to select the best 25 per cent of the group. If the selection is based on the individuals' true scores, we would choose the 25 individuals whose scores are 60 or more on the true score scale. These individuals are represented on the scattergram in the frequencies above the heavy

True Scores	Obtained Scores															f
	15	20	25	30	35	40	45	50	55	60	65	70	75	80		
60												1		1	2	
75												1	1		2	
70											1	1	1	1	4	
65									1	3	2	1			7	
60								2	2	4	2				10	
55								2	5	2	2				11	
50						1	4	5	3	1					14	
45					3	3	6	2							14	
40				1	4	2	1	3							11	
35				1	1	5	3								10	
30		1	2	2	2										7	
25			2	2											4	
20		1		1											2	
15	2														2	
f	2	2	4	7	10	11	14	14	11	10	7	4	2	2	100	

Fig. I. Possible distribution of frequencies of various combinations of obtained and true scores when the coefficient of correlation between them is .92. The scores along the top and left margins are the lower integral limits of the class intervals.

horizontal line. If true scores had not been available, and if the selection had been based on the obtained scores, we should have chosen the 25 individuals whose scores were 60 or more on the obtained score scale. These individuals are represented in the frequencies to the right of the heavy vertical line on the scattergram.

The two heavy lines, horizontal and vertical, divide the

scattergram into four quadrants. Consider the significance of the frequencies in each quadrant. The individuals whose scores fall in the upper right hand quadrant were selected on the basis of both true and obtained scores. These individuals, then, would have been properly classified on the basis of obtained scores alone. A similar statement applies to the seventy individuals whose scores fall in the lower left hand quadrant. They were properly classified, that is they were rejected, on the basis of obtained scores. But the ten individuals whose scores fall in the other two quadrants would have been wrongly classified on the basis of obtained scores alone. The five in the lower right hand quadrant would have been selected when they should have been rejected. The five in the upper left hand quadrant would have been rejected when they should have been selected.

In the preceding example it was assumed that the true score of each individual could be determined. This is impossible, of course, but the difficulty can be circumvented if the reliability coefficient of the obtained scores is known. For the square root of the reliability coefficient (the index of reliability) is the correlation between obtained and true scores.<sup>1</sup> Given this correlation, it is possible to use Pearson's (7) tables of volumes under the normal bi-variate correlation surface, or Thurstone's (11) computing diagrams for the tetrachoric correlation coefficient to predict the frequencies in each cell or quadrant of the scattergram.

Table 1 was prepared with the aid of Thurstone's diagrams. The figures in the body of the table are the expected percentages of incorrect classification. The table is entered with the values of the reliability coefficient of the test scores (on the left margin), and with the proportion to be selected or rejected (on the top margin). To illustrate the use of this table, consider the case in which 10 per cent of a group of 500 individuals have

---

<sup>1</sup> This statement is true if, and only if, the terms *true score* and *reliability coefficient* are defined consistently in any given instance. By *true score* we always mean the average of an infinite number of *similar* scores, but there are various possible definitions of *similar*. The definition chosen in any given instance depends on the reliability coefficient used in that instance. If, for example, an *equivalent forms* reliability coefficient is used, an individual's true score would be the average of his scores from an infinite number of such equivalent forms. If, on the other hand, a *test-retest* reliability coefficient is used, an individual's true score would be the average of his scores from an infinite number of testings with the same test.

been selected on the basis of test scores whose reliability is .90. The table shows that 4 per cent of this group (20 individuals) will probably be incorrectly classified. It would be expected that ten of these twenty will be wrongly included in the selected group. The other ten will be wrongly excluded.

## II. Errors in Marking

In school practice, numerical test scores are often converted into letter marks to facilitate interpretation, or to simplify record keeping. Again it is appropriate to ask how many marks

TABLE 1

*Estimated Percentages of Applicants for Selection Who Would Be Wrongly Classified on the Basis of Fallible Test Scores*

Coefficient of Reliability for the Test Scores	Proportion Selected (or Rejected)							
	.50	.40	.30	.25	.20	.15	10	.05
1.00	0	0	0	0	0	0	0	0
.96	5	4	4	4	3	3	2	1
.90	10	9	9	8	7	6	4	2
.85	13	12	11	10	9	7	5	3
.81	15	14	13	11	10	8	6	4
.76	17	16	15	13	12	10	7	4
.64	20	19	18	16	14	12	9	5
.49	26	24	22	20	18	15	11	6
.25	34	32	29	26	23	18	14	8
.00	50	48	44	38	32	26	18	9

will be wrongly assigned because of inaccuracy in the test scores. An extension of the methods just described will provide an answer to this question.

Let us consider first a situation in which scores from a single test are to be converted into letter marks. Suppose that a test is given to 1,000 individuals, that the obtained scores have a reliability coefficient of .90, and that, on the basis of the obtained scores, the best 5 per cent of the group are to receive the mark A, the next 25 per cent B, the middle 40 per cent C, the next 25 per cent D, and the poorest 5 per cent F. Knowing the reliability of the test scores, it is possible to determine the correlation between obtained and true scores, and from this to construct a scattergram which shows the relation between assigned marks and true marks. Either Pearson's tables or Thurstone's

diagrams may be used to obtain the various cell frequencies. Such a scattergram for test scores having a reliability coefficient of .90 is presented in Fig. II.

The frequencies in the five cells along the principal diagonal represent the 772 individuals who were assigned marks corresponding to their true marks. The remaining 228 individuals in cells off the principal diagonal were assigned incorrect marks. Of the 50 who were assigned A's, 12 should have received B's. Of the 250 who were assigned B's, 12 should have received A's and 45 should have received C's. Other columns in Fig. II may

True Marks	Assigned Marks				
	A	B	C	D	F
A	38	12	0	0	0
B	12	193	45	0	0
C	0	45	310	45	0
D	0	0	45	193	12
F	0	0	0	12	38

Fig. II. Frequency of error in assigning 1,000 marks on the basis of test scores whose reliability is .90.

be interpreted similarly. In all, approximately 23 per cent of the group were mis-marked.

Similar scattergrams were constructed for test scores having other coefficients of reliability. The percentage of mis-marking in each instance is given in Table 2. It will be observed that there is an appreciable percentage of mis-marking even when the score reliability is very high by ordinary standards. When the reliability is .75 or less, a small percentage will be mis-marked by two mark levels.

Let us consider next a situation in which letter marks are assigned, not on the basis of a single distribution of scores, but

rather on the basis of several such distributions. This occurs commonly in the assignment of final course marks. The probable frequency of incorrect marking in this situation may be estimated with the help of Table 2 provided (1) that the proportion of each mark assigned conforms to that on which the table is based (5 per cent A's, 25 per cent B's, etc.), (2) that the final marks are based on numerical composites obtained by adding numerical component scores and ratings, and (3) that sufficient data is available to permit an objective estimate of the reliability of the composite scores.

TABLE 2

*Estimated Percentages of Incorrect Marking Resulting from the Use of Fallible Test Scores When a Five-Category (5-25-40-25-5) Distribution Is Followed*

Reliability of Test Scores	Percentage of Incorrect Marking
1.00	0%
.99	5%
.98	9%
.95	15%
.90	23%
.80	33%
.70	40%
.50	50%
.00	70%

The reliability of composite scores depends essentially upon the reliability and variability of the component scores, and on the intercorrelations between them. If, as commonly happens, an instructor assigns final marks more or less subjectively on the basis of a variety of scores and ratings on tests, assignments, recitations, reports, etc., it is practically impossible to obtain an objective estimate of the reliability of the final marks.

Sometimes, however, an individual's final mark depends entirely on the sum of his scores on several unit tests. If these tests are sufficiently similar to justify the assumption that their reliability coefficients and score variances are the same, and that the intercorrelations between them are equal to the reliability coefficients, the Spearman-Brown formula may be used to estimate the reliability of the composite scores. If these assumptions are not justified, the computation is more involved.



A discussion of this problem, as well as formulae suited to various situations, are given by Jackson and Ferguson (4) in Chapter VI of *Studies on the Reliability of Tests*.

It should not be assumed that composite scores are always and necessarily more reliable than the most reliable of the component scores, for such is not the case. If it is desired to obtain maximum reliability in a composite of scores which vary in reliability and in their intercorrelations, the scores must be weighted.

### III. *Some Limitations*

The frequencies of incorrect selection and marking which have been presented here are estimates only. They indicate what may be expected, under given conditions, with large groups or in the long run. In any single instance, the actual frequency of error may be somewhat more or less than the estimate given. Further, the estimates were derived by assuming normality in the distributions of both obtained and true scores. It is unlikely, however, that the deviations from normality ordinarily encountered in score distributions are great enough to render the estimates inapplicable.

There are many sources of inaccuracy in test scores. Not all of these affect the reliability coefficient. Further, the various types of reliability coefficient (re-test, equivalent forms, consistency, etc.) are affected by different sources of score inaccuracy. The estimates here given include only those sources of error which affect the particular reliability coefficient used.

The basing of these estimates on reliability coefficients rather than on validity coefficients requires a word of explanation. We recognize that the validity of scores is more crucial than their reliability, and that many errors of classification can result from the use of highly reliable but invalid scores. The difficulty with using validity coefficients as a basis for estimates of incorrect classification is simply that dependable coefficients of validity are not easy to obtain, especially for tests of educational aptitude or achievement.

If validity coefficients are available, and if the proper corrections are applied to them, they may be used to enter the tables previously given to obtain an estimate of the frequency

of incorrect classification. There are three steps in the process. First, the validity coefficient must be corrected for attenuation due to the unreliability of the criterion measures (8, formula 122). Second, the corrected validity coefficient must be squared. When corrected and squared, the validity coefficient is analogous to a reliability coefficient with respect to our tables. Third, the squared, corrected validity coefficient is used in place of a reliability coefficient to enter the tables. The appropriate estimate of frequency of error in classification may then be read.

#### IV. *Related Studies*

A number of writers have dealt with the interpretation of test validity coefficients for specific purposes. In 1919, Kelley (6) discussed the classification of men, and suggested an inverse measure of the predictive value of a test, which he called the *coefficient of alienation*. This statistic measures the gain in precision of prediction which results from the use of a given test in a given situation. It indicated what fraction of the total group variability on the criterion remains among those who receive the same score on the predictor test.

Peters and Van Voorhis (8, pp. 508-510) give tables which show the probability of an individual reaching or exceeding a given decile standing on the criterion, given his decile standing on the predictor test, and its validity. Bittner and Wilder (1) provide a more detailed solution of this problem, and present formulae and tables showing the per cent of individuals making a given score on the predictor test who would be expected to equal or to exceed a given score on the criterion.

Taylor and Russell (10) have presented a discussion and tables of the proportion of individuals selected by a fallible test who will prove satisfactory in the job for which they have been selected. Their tables, like those presented here, were derived from the normal correlation surface. To use these tables it is necessary to know the validity coefficient of the selection test, the proportion of applicants to be selected, and the proportion of successful job-holders in the group on which the test was validated.

Richardson (9) has derived a formula for predicting the

increase in effectiveness of a group of personnel when the group is selected on the basis of scores from a test of known validity.

Several other reports on problems of selection do not involve the use of test validity coefficients. Johnson (5) has developed measures of the selective value of a test which he calls the "Coefficient of Selectivity," and the "Coefficient of Correctivity." These measures are based on the four-fold point coefficient of correlation between success on the test and success on the criterion. Burt (2) has called attention to a method for minimizing errors when test scores are used to select individuals who must meet a given standard on the criterion. He suggests that the pass mark on the test be determined by dividing the critical score on the criterion by the regression of the criterion on the test.

To improve the accuracy of selection without increasing the over-all amount of testing, Clarke (3) has recommended repeated testing on successively narrower border zones. A shortened first test is given to all applicants. This is followed by a second test for those who cannot be classified with confidence on the basis of the first test. The process may be repeated a third or fourth time if necessary.

It is evident that considerable thought has been given to the interpretation and use of test statistics in ways that have specific meaning. The present article represents one further effort in this direction. It is intended to provide answers to questions which arise frequently in the use of tests, but which have not previously been answered directly.

#### REFERENCES

1. Bittner, R. H. and Wilder, C. E. "Expectancy Tables: A Method of Interpreting Correlation Coefficients." *Journal of Experimental Education*, XIV (1946), 245-252.
2. Burt, Cyril. "Statistical Problems in the Evaluation of Army Tests." *Psychometrika*, IX (1944), 219-235.
3. Clarke, E. R. "Predictable Accuracy in Examinations." *The British Journal of Psychology Monograph Supplements*, No. XXIV. London: Cambridge University Press, 1940.
4. Jackson, Robert W. B. and Ferguson, George A. *Studies on the Reliability of Tests*, Bulletin No. 12. University of Toronto Press: Department of Educational Research, 1941.

5. Johnson, H. M. "A Useful Interpretation of Pearsonian  $r$  in  $2 \times 2$  Contingency Tables." *American Journal of Psychology*, LVII (1944), 236-242.
  6. Kelley, T. L. "Principles Underlying the Classification of Men." *Journal of Applied Psychology*, III (1919), 60-61.
  7. Pearson, Karl (ed.). *Tables for Statisticians and Biometricians*. London: Cambridge University Press, 1931.
  8. Peters, Charles C. and Van Voorhis, Walter R. *Statistical Procedures and their Mathematical Bases*. New York: McGraw Hill Book Company, 1940.
  9. Richardson, Marion W. "The Interpretation of a Test Validity Coefficient in Terms of Increased Efficiency of a Selected Group of Personnel." *Psychometrika*, IX (1944), 245-248.
  10. Taylor, H. C. and Russell, J. T. "The Relationship of Validity Coefficients to the Practical Effectiveness of Tests in Selection: Discussion and Tables." *Journal of Applied Psychology*, XXIII (1939), 565-578.
  11. Thurstone, L. L., Chessire, Leone and Saffir, Milton. *Computing Diagrams for the Tetrachoric Correlation Coefficient*. Chicago: University of Chicago Bookstore, 1933.
-

# THE VALIDITY OF STANDARD AND CUSTOM-BUILT PERSONALITY INVENTORIES IN A PILOT SELECTION PROGRAM<sup>1</sup>

DONALD E. SUPER

Teachers College, Columbia University

THE Aviation Psychology Program of the Office of the Air Surgeon, Army Air Forces, had, by the Fall of 1943, developed a battery of tests for pilot selection which had a validity of .66, and which incorporated in it most of the tests that were used throughout the rest of the war. This relatively high degree of validity not only made it possible for those associated with the selection and classification program to consider types of tests which had not been worked with during the early days of aviation psychology, but forced them into an attempt to find tests which not only had validity but which also had negligible correlations with other tests already in use. The stage had been reached at which a test with a validity of .20, and a correlation with the battery score, or stanine, as it was called, of .10 was more desirable than one with a validity of .40 and a stanine correlation of .50. As the battery included a number of valid tests of manual, spatial, and intellectual abilities, but only two tests which measured personality characteristics, it seemed that room for improvement in the battery lay primarily in the field of personality tests.

## *Early Personality Tests*

The two valid personality tests already in use were a *Biographical Data Blank* developed by Laurance F. Shaffer along the lines of one used by the Civilian Pilot Training Program

---

<sup>1</sup> Credit should be given to John V. McQuitty and William H. Angoff for statistical work associated with the projects reported herein; to Seymour P. Stein and associates for administering the tests; and especially to Robert R. Blake, John L. Wallen, and Joseph Weitz for creative participation in the development of the *Satisfaction Test*.

and by the Navy (validity = .33), and a *General Information Test* to which Nicholas Hobbs, Robert R. Blake, John C. Flanagan, Frederic B. Davis, and the writer contributed at different times and in different ways (validity = .51). The titles of both tests are descriptive; in both cases scoring was by means of empirically developed keys, which gave credit for having done or for knowing things characteristically done or known by successful fliers. Other tests designed to measure personality had also been tried, but like the above they were not subjective in the same sense as personality inventories such as the Bernreuter. Instead, they attempted to measure personality as manifested by steadiness in a manual task despite subjection to emotional pressure in the form of noises, "razzing," etc., or as shown in resistance to distraction while performing routine computations. Neither of these approaches proved valid for aircrew. Work with personality inventories had been viewed with disfavor, because they seemed likely to be too much affected by a desire to make a good impression in a selection situation in which the opportunity to fly and to be an officer was at stake. As a result, only one test of this sort had been experimented with prior to the Summer of 1943. This was the *Satisfaction Test*, developed by the writer with the assistance of Robert R. Blake, John L. Wallen, and Joseph Weitz, described in some detail in the body of this paper.

### *Validity of Clinical Tests*

With the need to explore more fully the more baffling and less promising measures of personality, and with the stability of a program which had proved its value with less questionable techniques, a variety of personality measures were tried after the Summer of 1943. Some of these were so-called clinical techniques, such as the Rorschach, observation in semi-structured situations, and standardized interviews. It is not intended here to discuss these techniques in detail, but the general nature of their results should be noted in passing, for comparison with the results obtained with other instruments. The clinical techniques were applied to 660 aviation cadets in June, 1943, data for from 250 to 350 being used in validating the various tests

and ratings. The following summary is taken from an official report:

- a. Rorschach Test.
  - (1) Overall ratings: low positive validity of doubtful significance.
  - (2) Single categories: low validity for scattered scores, of doubtful significance.
  - (3) Pattern score: no validity.
- b. Interaction Test (ratings of behavior in group administration of three Wiggly Blocks to four cadets).
  - (1) Overall ratings: low positive validity of doubtful significance.
  - (2) Trait ratings: no validity.
- c. Observational Stress Test (ratings of behavior in responding to confusing multiplicity of stimuli while taking a miniature-situation test).
  - (1) Overall ratings: low positive validity of doubtful significance.
- d. Observation During Psychomotor Test Rest Period (in small room, among selected aviation objects including a bomb, a piece of broken and twisted fuselage, etc.)
  - (1) Overall ratings: no validity.
- e. Interview.
  - (1) Overall ratings: no validity.

Some of these techniques might, it should be remembered, prove valid for the selection for other types of assignment or for clinical diagnosis. But the studies just referred to demonstrated that, as used in these experiments, they had no real validity for selecting Army pilots.

#### *Validity of Standard Personality Inventories*

The published personality inventories administered at Psychological Research Unit No. 1 during 1943 and early 1944 included the *Shipley Personal Inventory*, the *Adams-Lepley Personal Audit*, the *Guilford Inventories*, the *Maller-Glaser Interest-Values Inventory*, the *Humm-Wadsworth Temperament Test*, and the *Minnesota Multiphasic Personality Inventory*. Because of the tapering-off of test development for aviation cadet classification after the Spring of 1944 and the shifting of personnel to research and service in training, combat, redistribution, and rehabilitation, data for some of these tests have still not been analyzed. Data for the Shipley, Adams-

Lepley, and Humm-Wadsworth are available, and are reported in summary herewith:

- a. *Shipley Personal Inventory* ( $N = 561$ )
  - (1) Biserial correlation with graduation-elimination in primary pilot training:  $-.12$ .
- b. *Adams-Lepley Personal Audit* ( $N = 271$ )
  - (1) Biserial correlations between the nine scores and graduation-elimination in primary pilot training ranged from  $-.12$  to  $.06$ .
  - (2) Biserial correlation between overall rating of adjustment (a clinically derived pattern score) and graduation-elimination:  $-.08$ .
- c. *Humm-Wadsworth Temperament Test* ( $N = 202$ )
  - (1) Biserial correlation between Humm-Wadsworth scores and graduation-elimination in primary pilot training include coefficients of  $-.72$  (epileptoid),  $-.18$  (hysteroid), and  $.16$  (autistic), but the other coefficients range from  $-.01$  to  $.05$ .

In all cases in which relationships are known, e.g., for the *Shipley Personal Inventory*, the correlation between inventory and battery is negligible. Only in the case of the Humm-Wadsworth epileptoid and hysteroid scales are the validity coefficients high enough to suggest genuine relationship, but the numbers in this case are not large enough for them to be significant. The conclusion to be drawn is, then, that these three standard personality inventories added nothing to the predictive value of the battery, having no significant relationship with success in flying training.

#### *Construction of a Custom-Built Personality Inventory*

It is interesting, in view of the lack of validity of other personality measures, and especially of standard personality inventories, to turn to the data concerning the custom-built personality inventory referred to earlier, the *Satisfaction Test*. In order that the implications may be clearly brought out, its development will first be described, after which the data on its validity will be presented.

In August, 1942, two questionnaires were prepared and administered to approximately 300 cadets each at Psychological Research Unit No. 1. In the first they were given two blank



sheets of paper, on which they were asked anonymously to list, on one sheet, all of the things they liked and, on the other, all of the things they disliked about Army life. Five minutes were allowed for the likes, ten for the dislikes. The second questionnaire, administered to another group of cadets on the next day, consisted of paired comparisons, Part I covering things which the test constructors thought might be liked by soldiers, cadets, and fliers, and Part II covering things which it was thought they might dislike. Traits used in developing the paired comparisons included such things as the presumed extraversion of the effective combat pilot, the supposedly scientific and sedentary interests of the navigator, and the assumed desirability of good morale in any type of assignment. The responses to these two questionnaires were then analyzed in order to provide realistic choices and to couch them in terms used by cadets.

The *Satisfaction Test*, Form A, was developed from this material. It consisted of four parts with a total of 150 items. Part I, Cadet Likes, consisted of paired comparisons such as:

As a cadet I would get more satisfaction from

1. A. the chance to travel more.  
B. the good pay.
2. A. the orderliness of life in the Army.  
B. the association with other fellows.

Part II, Cadet Dislikes, similarly consisted of paired comparisons concerning irritating or unsatisfactory aspects of Army life. Part III, Cadet Opinions, dealt with attitudes toward situations and activities encountered. These were multiple-choice items such as:

83. Stunt flying
  - A. should give you a great thrill.
  - B. ought to be prohibited.
  - C. is not as much fun as flying to some objective.
  - D. frequently causes passengers to become sick.
  - E. interests me less than straight and level flying.

Part IV, Preferences, consisted of civilian sports and leisure-time activities believed by the test constructors to be related to personality traits. Time limits were established for each part in such a way as to permit most cadets to finish. The test was administered to 2000 cadets as a routine part of the classifica-

tion test battery, and answer sheets were filed to await the receipt of criterion data in the form of pass-fail reports from primary flying schools.

In the meantime two a priori keys were developed. One of these was constructed solely on the basis of clinical judgment by a member of the unit with appropriate background, based on the significance of each possible response for good or bad morale: this classification was checked by the others working on the test, so that the key represented the judgments of four psychologists. The second a priori key was more empirical, in that it was based on the atypical answers of tested cadets. Any response chosen by less than 20 per cent of the responding group in the case of paired comparisons, or by less than 10 per cent in the case of multiple-choice items, was included in the *atypicality* key.

When the criterion data were received two other keys were constructed, both entirely empirical in nature. Answer sheets which had been completed by cadets with testing numbers ending in an odd number were item analyzed. Data were available for 460 successful and 327 unsuccessful cadet pilots with odd testing numbers. All items which showed a difference between passes and failures which was significant at the 10 per cent level or better were critically examined from the point of view of logic. On this basis a 75-item key was developed. The even-numbered cases ( $N = 460$  successes and 338 failures) were treated in the same manner, with a resulting 56-item key.

#### *Validity of the Custom-Built Personality Inventory*

After the odds and evens keys had been constructed on the basis of item validities, the answer sheets were scored with the four keys: morale, atypicality, odds, evens. In the case of the first two keys, the even-numbered cases were used for validation purposes (either group could have been used); in the case of the odds key, even-numbered cases were used for validation; and in the case of the evens key, odd-numbered cases were used for validation. The results appear in Table 1.

The clinically derived or a priori morale key had no validity for success in flying training. The atypicality key correlated

negatively with success, but the coefficient was too low to be significant. The two strictly empirical keys, based on differences between successful and unsuccessful cadet pilots, had low but reliable validities of .18 and .21. By commonly accepted standards these are low, but as the two keys had negligible correlations with the selection battery, their use would actually increase the validity of a battery with a multiple-validity coefficient of .50 by .02 or .03. This is an appreciable increase, worth the time required in testing and scoring when the battery in use has already reached a point of development ( $R = .66$ ) at which the obtaining of higher validities is extremely difficult. As other valid apparatus and paper and pencil tests then being

TABLE 1  
*Validity of the A Priori and Empirical Keys for the Satisfaction Test  
in Pilot Selection (Primary Flying)*

Key	Group	N Successes	N Failures	$r_{bis}$	$r_{battery}$	Add to R of .50
Morale	Evens	460	327	-.01	x	x
Atypicality	"	460	327	-.12	x	x
Odds	Evens	460	327	.18	.12	.02
Evens	Odds	460	338	.21	.04	.03

developed generally correlated highly with the battery, this test was a valuable addition to the battery. It contributed more than standard personality inventories and clinical tests which also measured different characteristics but which had no predictive value.

#### *Revision and Revalidation*

Despite the conclusions just drawn, the *Satisfaction Test* was not added to the standard battery. Some of the more laboratory-minded aviation psychologists still harbored prejudices against paper-and-pencil tests which were not dispelled until later studies showed that they could do all that apparatus tests could at the expense of less time and money. Some of the applied psychologists still looked askance at "subjective" tests despite the evidence presented, and probably no one as yet fully appreciated the value of a validity coefficient of .20 combined with a battery correlation of .10. It was therefore decided to

revise the *Satisfaction Test* in order to try to raise its validity, to shorten the required testing time, and to demonstrate its continued validity with other groups.

Accordingly the valid items of Form A were studied and grouped according to content, in order to obtain a personality picture of the successful cadet pilot. The following sketch was written on the basis of this logical analysis of the empirically valid test items:

The successful cadet pilot is an active, aggressive individual who dislikes slow-moving situations. He wants to learn, but would like to learn in an active rather than in a passive or receptive way. He likes security of the kind the Army offers, that is, having his meals, lodging, medical care, and the like taken care of automatically, but he likes also the variety, action, and thrills of Army life and of flying. He is realistic and concrete in his attitudes toward the war rather than idealistically patriotic. He faces difficult situations directly rather than dodging them, likes to know the facts, and chooses action rather than inaction. He likes authority and strives to attain it, but he accepts the authority exercised by others. He is impatient with details. He likes fun. He is not sadistic, but has no great love for others, is no humanist. He is challenged by the power of machines and wants to rule them.

With this statement and related material as guides, statistically valid or nearly valid items were selected from Form A, and new items were constructed. When it seemed advisable for clarity or style, old items were modified slightly.

The revised test contained 85 items in two parts. Part I consisted of 25 items, each item having two possible answers, as follows:

5. Getting out of step while marching makes me
  - A. angry and determined to try harder to keep in step.
  - B. embarrassed at being so conspicuous.
6. As an aerial gunner, I would get more satisfaction from
  - A. matching my skill against the enemy.
  - B. seeing the enemy go down in flames.

Part II consisted of 60 paired comparisons such as the following:

If given the choice and having equal opportunity and ability, would you rather

28. A. design a new plane?  
B. test a new plane?
53. A. ambush the enemy?  
B. storm an enemy position?

This test was prepared in two Forms, B and C, identical except that B provided only two alternatives, forcing a choice, whereas C contained a third choice for each item which permitted the respondent to indicate that neither answer was appropriate for him. A time limit of twenty minutes was found suitable, and the two forms were administered on alternate days in September, 1943, to more than 2000 cadets each, at Psychological Research Unit No. 1.

After the usual lapse of time criterion data became available, and empirical keys were developed according to the standard method described above. The new validities, like the old, were about .20, but the testing time had been reduced by one-half, the correlation with the pilot stanine was as low as before, and the validation group being more restricted in range as a result of higher selection standards, the predictive value of the test was actually higher than the statistics indicate. Then, too, the value of low validities in unique tests was now better recognized. It was therefore recommended by Headquarters, AAF Training Command, that the *Satisfaction Test* be incorporated in the classification battery. By this time, however, it was Summer, 1944. There were reasons for keeping the battery in its existing form rather than continuing to make changes. There were still misgivings in some places concerning subjectively answered tests. The recommended action was therefore never actually taken.

### *Conclusions and Implications for Civilian Work*

Experience in the use of clinical tests, standard inventories, and custom-built inventories of personality in the selection of Army aviation cadets for pilot training has been reviewed, and the following findings reported:

1. Ratings of probable success based on clinical techniques were not useful in selecting pilot trainees.
2. Scores obtained from clinical tests such as the Rorschach were not useful in selecting pilot trainees.
3. Scores obtained from standard personality inventories such as the *Shipley Personal Inventory*, the *Adams-Lepley Personal Audit*, and the *Humm-Wadsworth Temperament Test* were not useful in selecting pilot trainees.

4. A priori keys (one based on clinical judgment of significance for morale and one based on empirically determined atypicality) for a custom-built personality inventory were not useful in selecting pilot trainees.

5. Empirically derived keys (based on success-failure) for a custom-built personality inventory, in which an attempt was made to tap behavior in situations meaningful to aviation cadets, proved to have sufficient value in selecting pilot trainees for the test to be recommended for inclusion in the classification battery.

The following principles applicable to civilian work with personality inventories in vocational selection and guidance are believed to emerge from the military experiences reported in this paper:

1. When a valid battery of aptitude tests has been developed and new aptitude tests are found merely to measure the same thing in different ways, thereby adding little to the validity of the existing battery, personality inventories may be worth considering.

2. In such a situation, the personality inventory may have low validity, both absolutely and relatively to the aptitude tests, but, if the relationship to the criterion is significant, it will have a unique contribution to make to the battery.

3. Standard personality inventories are less likely to be valid, because of their general terms and situations, than custom-built inventories based on analyses of the behavior and attitude-evoking situations in the vocation or in the employing organization.

4. Empirically validated success-failure keys, checked against the logic of the situation and of the item, are likely to prove more valid than keys based on clinical judgment or on an internal rather than external index of validity.

## THE GRADUATE RECORD EXAMINATIONS<sup>1</sup>

K W VAUGHN

Director, The Graduate Record Examination

THE *Graduate Record Examination* Project of the Carnegie Foundation for the Advancement of Teaching was inaugurated in 1936. After two years of experimentation and research, a final form of the Examination was produced. Last year, two new types of tests were introduced into the series. Therefore, the title of this paper is no misprint; there are now three *Graduate Record Examinations*.

### I.

As originally conceived, the *Graduate Record Examination* was appropriately named. The first examination was designed to provide objective measures of the college graduate's scholastic record with respect to knowledge. Initiated by the graduate deans of Harvard, Yale, Princeton, and Columbia, the examination was constructed in collaboration with the undergraduate and graduate faculties of these universities. As a supplement to and substitute for the college record, the Examination was used as a record of the graduate's scholastic preparation. In interpreting the word "graduate," however, it must be explained that the Examination was designed for the purposes of the graduate schools; consequently, the tests were adjusted to the abilities of first-year graduate students, and the content and emphasis of the tests were influenced, if not determined, by graduate faculties. These facts are well known to most of you, yet they must be understood and appreciated in relation to the development of the new types of tests which were constructed.

The project was admittedly an experiment—on the part of

---

<sup>1</sup> A paper given at the meetings of the American College Personnel Association in Columbus, Ohio, March 29, 1947.

the graduate schools as well as on the part of the Foundation. While the Examination Staff did not publish systematic studies of the results of the early experimentation, the cooperating graduate schools did conduct studies and the results were published under the auspices of the project. In general, the results obtained tended to inspire confidence in the use of the examinations for the selection of graduate and professional school students. As other graduate schools cooperated in continued experimentation, this confidence met no serious reversals; it tended to increase as graduate faculties had an opportunity to study the results for their institution.

Naturally it was desired to prepare as appropriate and precise a series of tests as could be developed. For the small group of institutions which originally cooperated in the project, the Examination was particularly appropriate and carefully constructed. In this sense, then, the Examination was as well designed and carried out as could be expected in view of the approach to measurement and the scope of experimentation.

Without this emphasis upon quality of content and careful preparation, it is doubtful whether the experiment could have succeeded. It is questionable, also, whether the deans of graduate and professional schools, and their faculties, which later reviewed the test materials, would have placed in them as much confidence as was true during the period when the experiment was in its crucial stages. Yet in spite of this careful construction of the Examination, the approach to measurement decreed that the tests were not universally appropriate. By restricting the emphasis of the tests to knowledge alone, the curricular peculiarities of eastern liberal arts colleges were emphasized. Similar criticisms, mostly legitimate, were made.

For a time it seemed that the conclusions of this experiment on any wide scale would be negative—positive only in those colleges whose curricular offerings paralleled those of the original cooperating universities. The practice of measurement of the attainments of advanced college students by standardized tests received little or no wide-scale impetus before 1943. During the war, of course, graduate and professional schools were decimated. There were few students to be selected; diminished



graduate faculties did not inquire systematically into the previous preparation of all applicants. Those graduate and professional schools which did make a serious effort to maintain standards during this difficult period, however, requested and effected the system of Examination Centers which now serve the Graduate Record Office.

What seemed a failure in 1944 was in 1945 a budding success. At this vantage point two years later, the measurement of educational outcomes represents a movement in American higher education which appears to offer greater assistance in the solution of student personnel problems than at any time in the past. It is perhaps of more significance that college and university faculties are not only willing, but in many cases eager, to obtain more dependable measures of the success of individual students than are provided by teachers' marks. In this sense, at least, the *Graduate Record Examination* experiment has been successful.

## II.

In practice, *Graduate Record Examination* is a misnomer. Very early in the experiment, the Examination was administered widely to senior students in colleges of liberal arts for purposes of evaluating the achievement of individuals and groups. Later a number of colleges requested that it become available for use with sophomore classes for purposes of measuring growth and development in general education during the last two years of college study. Thus new purposes were introduced; the original series of examinations remained unchanged. The Examination used in this manner could no longer be considered strictly as a substitute for the college record and the tests provided, constructed for one purpose, were being accepted and used for other purposes. I do not suggest that the test was seriously misused, but rather that it was used for purposes for which it was not particularly designed and for which it lacked the type of precision that is most desirable in any examination. In the absence of other standardized tests, the first form of the Examination offered at least one basis of evaluating the status and progress of classes of undergraduate students. While no serious harm was done, undergraduate colleges using the Exami-

nation for purposes of evaluating senior classes were using an instrument which emphasized the educational values held by graduate schools. I suggest that the aims of undergraduate education include for certain students those fostered by the graduate and professional schools, but that, for some students at least, these goals do not represent a "fair" basis for evaluating the outcomes of a college education.

As in the case of all experiments, the *Graduate Record Examination* project must sometime be ended. More than two years ago, the Carnegie Foundation and Carnegie Corporation began to collect data and opinion as to the success of its experiment. One aspect of the matter is concrete. To obtain substantial evidence of the educational climate with respect to the use of the Examination, the fee, formerly a nominal one, was increased to the point where with extensive use the routine services of the examinations office could become virtually self-supporting. It was believed that, if the Examination served a useful purpose, colleges and universities would be willing to pay for the services at a rate which would insure the continuance of the service and not require continued subsidy on the part of Carnegie Corporation. The success of the Examination under these conditions is most promising.

Most of you, of course, are aware of the fact that Dr. O. C. Carmichael, President of the Carnegie Foundation, has appointed a committee to review the merits and limitations of a consolidation of the principal non-profit testing agencies. Whatever the results of these deliberations, at least one fact now appears to be clear. Colleges and universities, graduate and professional schools, have come to realize and depend upon educational measurement as a useful tool, one whose possibilities on a national scale are still far from being realized.

### III.

Before discussing the two new types of tests of the *Graduate Record Examination* series, let us consider first certain guiding principles which should constitute the frame of reference for the operations of any national examination service.

1. A national examination agency should serve one or more definite purposes. It must meet at least one and prefer-

- ably several needs in American education which are fairly universal.
2. Examinations provided by the agency must be of the highest quality. These instruments should be so designed to fulfill their function in a most efficient and dependable manner.
  3. The agency must command sufficient respect and confidence on the part of the scholars and leaders in the educational areas served in order to obtain the best cooperation, criticism, and advice on the development of tests and in the consideration of educational problems.
  4. The examination service must maintain continuous investigations of the area of education it serves, the effectiveness of its own tests for the purposes for which they are designed, and the extension of measurement techniques to related educational areas.
  5. The examination service must provide accurate, prompt, and efficient service where centralized scoring and reporting services are required. These must be accomplished in the most accurate manner and in the least possible time.
  6. The service must make provision for giving competent professional advice on measurement problems with respect to the tests it provides, and it must be prepared to describe their possibilities and limitations as compared with tests which are intended to serve the same or similar purposes.
  7. The examination service should provide a central agency which serves as an official clearing house of the records of individual students and as a resource for researches into educational problems.
  8. The examinations produced by the agency should be so designed and safeguarded that their contents and emphasis will not determine the direction of educational objectives in schools and colleges. National achievement examinations should be patterned after existing curricula and used to ascertain whether or not expressed goals are being achieved satisfactorily on the part of individuals, schools, and groups of educational institutions.
  9. The examination program must be available to educational institutions and to individuals at a reasonable cost. Certainly, the cost involved should not prohibit participation on the part of institutions or eliminate deserving individuals from consideration. The fees, however, must provide sufficient funds to enable the examination service to accomplish all of its aims effectively.
  10. Finally, the examination agency should be controlled by and its policies determined by an impartial board of recognized educators representing the educational several areas served by the agency.

In keeping with the purpose of this paper, only the first two of these criteria will be discussed further, and I shall deal only with certain major considerations involved in these principles.

#### IV.

There is abundant evidence of several needs in higher education for a national examination service which can provide examinations designed to fulfill these functions: (1) to aid in the selection of graduate and professional school candidates; (2) to provide a more objective basis of evaluating the outcomes of general education at the college level; and (3) to provide a basis for determining the extent and nature of the growth and development of the student's general educational background and achievement as he progresses through the college.

Some evidence of the need for examinations at this level is reflected in the number of students tested by the Graduate Record Office during the past two and one-half years. In 1944-45, the Examination was administered to 6,446 advanced college students. In 1946, a total of 23,143 students took the Examination. During the present fiscal year, ending July 1 next, it is estimated that approximately 50,000 students will take the Examination. Estimates in this respect may be conservative. One of the most recent orders of test materials from a single university was for 3,000 copies. One institution will test in a university-wide program involving only seniors and sophomore students approximately one-half as many students as were tested during the entire year of 1944-45.

University examination services are, of course, developing rapidly. The university which has sufficient funds stands in a position to be able to determine for itself certain aspects of its own effectiveness much more satisfactorily than is possible through a national examination service. It is interesting to note, however, that some of the strongest advocates of the *Graduate Record Examinations* are to be found in universities which have a well-developed examination service. The preparation of a standardized examination suitable for the purposes of measuring the outcomes of college education is an extraordinarily expensive and time-consuming task. Furthermore, few

universities are in a position to conduct national testing programs to prepare norms and determine educational levels. It appears, therefore, that national testing agencies such as the College Entrance Examination Board, the Graduate Record Office, and the Cooperative Test Service of the American Council on Education do serve needs in higher education and will be successful in proportion to the extent to which they are able to conform to the principles outlined above.

## V.

The essential features of any examination service are, of course, the tests and examinations provided. Without good examinations as the principal capital, no examination service can render to education its due, even though the agency may prosper financially. In my opinion, it is the first obligation of the professional staff of an examination service to insure high quality in its examination product. For this reason, the staff of the Graduate Record Office is the greatest critic of its own examinations. Because of shortcomings in our tests with which we are best familiar, we are making efforts to improve each test in the Graduate Record Examination series.

While it is nowhere expressed concretely in the record, the purpose of the original *Graduate Record Examination* was undoubtedly that of providing a basis of predicting success in graduate and professional schools. To accomplish this purpose, a series of tests in broad areas of subject matter, comprising the principal fields of undergraduate study, and a series of advanced tests in specialized areas, were constructed. The assumption underlying this basic plan is undoubtedly sound. It was adopted in light of the facts that the undergraduate college record often presents little or no possibility of obtaining comparable measures of the student's status at the end of college, that his marks were not dependable, and that test results might provide a more reliable measure of the student's general development and of his achievement in the area of specialization than would otherwise be available.

In measurement problems concerned with the future success of students, we are sometimes not as much concerned with look-

ing backward as we are with looking forward. That is, however reliable the student's previous college preparation may be, this preparation may represent general maturity which is necessary to advanced study. It may not reveal the specifics which are most essential in graduate study. Therefore, it seems reasonable to suppose that it is possible to construct a shorter examination whose purpose is to predict success in the graduate and professional school more effectively. We may be able, for example, to disregard subject fields and to prepare a composite test which would be made up of questions measuring those particular abilities which are most crucial in graduate study.

If, on the other hand, we are concerned with the measurement of educational growth and development in the undergraduate college, and recognize the fact that a large proportion of college students do not expect to go on to advanced study, then we must provide for measuring outcomes of college study which are not necessarily those deemed most important by the graduate school.

We have thus arrived at the point where we have two distinct general purposes and, since these purposes may not be identical from the measurement standpoint, the Graduate Record Office has constructed a series of *Tests of General Education* and a *Graduate Aptitude Test*. These tests are designed to be used in conjunction with the Advanced Test series of achievement examinations in specialized fields of study.

Measuring the outcomes of general education at the college level is a very difficult problem. In the first place, there are as yet few well-defined patterns of undergraduate study previous to the period of specialization, and furthermore, where programs are well developed there is often little agreement. Thirdly, one must rely almost entirely upon expert opinion for criteria of validity.

It is believed that programs of general education are of such diverse character that the ordinary type of achievement test will not produce satisfactory measurement in a national testing program. This was our principal criticism of the original form of the examination as applied in all types of undergraduate colleges. Based primarily upon subject matter, the traditional

achievement examination produces weird results—too frequently inexplicable. A general achievement test for use in a national testing program either at the college or secondary-school level finds this approach inadequate. A second approach which has been employed in a limited number of achievement examinations is that of organizing the tests around the objectives of instruction in the field without systematic regard to the subject matter presented for the purpose of realizing these objectives. The new *Tests of General Education* in the series employ a modification of each of these approaches to the measurement of general achievement, that is, both objectives of instruction and the general areas of subject matter common to them are considered.

By dealing with fundamental problems the tests in this series are intended to measure general achievement. The principal emphasis is concerned with the individual's ability to comprehend exactly, to think clearly in terms of specific problems and ideas, and to evaluate critically. These tests are free, for the most part, of the demand for remote and detailed information.

This does not mean that informational background or knowledge is unimportant or that it has not been systematically sampled in the *Tests of General Education*. These tests hold the student responsible for a very extensive and substantial background in the various fields of knowledge. The student is required to draw upon this background repeatedly but the knowledge required is concerned with important concepts, principles, and ideas rather than with isolated and descriptive facts. Furthermore, it is recognized that students acquire an important fraction of their knowledge and ability to deal with knowledge on their own motivation quite apart from formal class work. These tests, therefore, seek to derive an index of achievement which reflects not only the personal differential acquired in formal study, but also the increments that arise from intelligent observation, discriminating general reading, and other methods of self-education as well.

By concerning itself with knowledge exclusively and testing for the recall of precise information, the original form of the

Examination did not provide adequate measures of the ultimate objectives of college study. The college educator is ordinarily less concerned with a multitude of details than he is with the indispensable facts and relationships in the structure of knowledge. He seeks to develop in students the power to grasp important concepts and to employ trustworthy methods of work. He looks for ability to understand the meaning and significance of new situations, to think through new problems, and to evaluate them critically. In measuring the ultimate objectives of the educational process at any level, the tests employed must provide ample opportunity for the student to demonstrate his ability to use the knowledge he has acquired. In this respect the new *Tests of General Education* represent an important improvement over the first series of Profile Tests.

In the opinion of the cooperating committees this series of tests can be used for the following purposes:

- (1) to appraise students' achievement in broad general areas of study;
- (2) to provide a more adequate basis for the educational and vocational guidance of sophomore, junior, and senior college students;
- (3) to assist in the admission to and the appropriate placement in a program of general education of those students who transfer from one institution to another;
- (4) to assist in adapting instruction in general education to particular student needs;
- (5) to provide a trustworthy basis for studies of educational growth and maturation as students proceed through the college; and
- (6) to provide an important basis whereby the institution may appraise in a general manner its own effectiveness in relation to established goals.

It cannot reasonably be expected, however, that a series of tests designed to meet these needs in higher education can also be expected to predict success in the graduate school in the most efficient and precise manner.



## VI.

It may be possible to accomplish the second general purpose of the *Graduate Record Examination*—that of predicting success in graduate and professional school study—efficiently and effectively by a highly reliable test of general ability used in conjunction with a searching achievement test in the field of specialized study. Certainly, the total amount of testing time required can be reduced materially, and a vigorous policy of item selection can produce unusually high reliability.

Developed during 1945-46, the *Graduate and Professional Aptitude Test* is intended to provide differential measures of general scholastic ability. This test was first administered on a trial basis to approximately 2,000 first-semester graduate students, and, on the basis of this administration, a second form was constructed. This revised form was used first on a national scale in the Association of American Medical Colleges testing program on January 11 of this year.

The general scholastic ability section of the *Graduate Aptitude Test* includes two tests—*Verbal Ability* and *Quantitative Ability*. Six scores are derived from this section: four verbal scores—Scientific, Social, Humanistic, and Composite; one quantitative score; and an Index of General Ability. These scores are to be used in conjunction with the score obtained from the *Advanced Test* selected by the student.

The *Graduate Aptitude Test* is designed particularly for prospective graduate and professional school applicants, and for matriculants. The scaled-score system provided with this test will be based on the performance of first-year graduate students admitted to candidacy for an advanced degree in a carefully chosen, representative sample of graduate colleges and professional schools.

Although the *Graduate Aptitude Test* will be available in final form on August 1, 1947, its use will still be experimental in character. The further refinement of the tests will be based on researches conducted in a number of cooperating graduate and professional schools—researches which will be designed to determine its value for predicting the academic success of various

groups of graduate students. These researches will serve also as a basis for further development and refinement of the test.

## VII.

In conclusion, I should like to emphasize that we maintain a practical attitude toward the tests developed in the *Graduate Record Examination* series. We are approaching all of our problems of test improvement and refinement by research methods and as rapidly as time and funds permit. We feel obligated, however, to supply our services and assist with the many problems now confronting all phases of higher education. We are dissatisfied with our tests, and this dissatisfaction, we believe, will result in their improvement.

Finally, the Graduate Record Office represents a conservative position with respect to the use of test results. However well refined, however carefully validated, tests are but one tool with which the educator works. All of the *Graduate Record Examinations*, and any of the other tests prepared and used in our testing projects, can be of some assistance at various stages in higher education. None of these tests, however, can reduce the task of student appraisal to a routine business.

•

# VALIDATION OF THE STUDY OF VALUES FOR TWO VOCATIONAL GROUPS AT THE COLLEGE LEVEL

HAROLD G. SEASHORE<sup>1</sup>  
The Psychological Corporation

## *Purpose and Source of Data*

EDUCATIONAL and vocational advisors in professional and liberal arts colleges often use the Allport-Vernon *Study of Values* (1, 5) for appraising the interests of their students. The six "values" (which represent Spranger's "type of men") measured by this questionnaire-type instrument are said to indicate broad directions of the person's fundamental motivations. Research studies (3) tend to support the predictions regarding the kinds of profiles that given groups of persons should produce. Although it deals only in broad categories of interests, the *Study of Values* has been found useful in counseling with students and other adults regarding their more precise occupational plans.

This paper presents data concerning two rather well-defined college groups. The sample consists of 459 men majoring in Health and Physical Education (referred to as H & PE in the following pages) in Springfield College. The second sample is of 252 men majoring in Applied Social Sciences in the same college. Eighty-nine per cent of the men filled out their questionnaires during their freshmen or sophomore year; the rest were upperclassmen.

Before collecting these data the author studied the Spranger definitions, as presented by Allport and Vernon, and also studied occupational activities contemplated by the students in the two curricula. On the basis of this study predictions were made regarding the pattern of interests which the "typical major" should show.

<sup>1</sup> The research was carried out while the author was a member of the Faculty of Springfield College

It was anticipated that the average H & PE major would have a peak in his profile in the *political* category since a common motivation of these men is competitiveness in group and individual sports. Spranger's political type included persons striving for power, whether in politics, law, warfare, or sports. A secondary peak in the profile was expected in the *social* field since this group of men was preparing to become teachers and leaders in physical education in schools, social agencies, com-

TABLE 1  
*Profiles of Two Springfield Groups and Three Other Groups  
on the Study of Values*

Groups		Theo- retical	Eco- nomic	Aes- thetic	Social	Politi- cal	Reli- gious
1. 459 Men, Majors in Health and Physical Education	Mean	30.30	30.24	23.15	30.85	32.80*	32.74
	SD	6.21	6.18	6.44	5.62	5.95	7.11
2. 252 Men, Majors in Applied Social Sciences	Mean	30.87	29.08	24.44	32.03	28.88	35.19
	SD	6.22	6.48	6.93	5.43	6.51	7.56
3. 279 Dartmouth Students (Stone)	Mean	29.67	32.14	30.37	30.18	31.99	25.66
	SD	7.72	9.69	9.35	5.54	7.16	9.19
4. 160 Men, Majors in Health and Physical Educa- tion (Brown)	Mean	31.22	31.15	22.86	32.58	33.55	28.76
	SD	Not reported					
5. 1163 Adults (Cantril and Allport)	Mean	30.83	32.02	27.04	29.74	32.08	27.96
	SD	Not reported					

\* The two highest scores in each profile are printed in italics.

munity recreation centers, and similar socially oriented organizations serving youth. The Springfield H & PE major, it was believed, would differ from majors in this field in general in two ways. First, a large proportion of these men planned careers in the non-coaching aspects of physical and health education. Second, the selection policy of the college favored the admission of men with all-around ability rather than men with outstanding prowess in one or two fields. Because the college has a religious emphasis, however liberal, it was felt that the *religious* motive might appear as the third dominant interest, with the possibility that the *theoretical* scores might exceed it. It is

considered that the *economic* and *aesthetic* motives should be quite subordinate.

The expectation for the Social Science majors was that their highest points would be *social* and *religious*, with *theoretical* as a third positive factor.

### Results

Table 1 presents the data for the two Springfield College samples and also comparative data from three other studies by Cantril and Allport (3), Stone (4) and Brown (2). Table 2

TABLE 2

*Critical Ratios Indicating Significance of Differences between Average Scores of Stone, Springfield H & PE, and Springfield Social Science Groups\**

	Theo- retical	Eco- nomic	Aesthetic	Social	Political	Religious
Springfield H & PE						
Stone Sample	1.15	-2.92	-11.28	1.60	1.59	11.06
Springfield Soc. Sci.						
Stone Sample	1.97	-4.31	-8.35	3.85	-5.27	13.05
Springfield H & PE						
Springfield Soc. Sci.	-1.16	2.32	-2.43	-2.74	7.84	-4.22

\* The italicized CR's are positive, meaning that the first group of the two being compared has a *higher* average score.

shows the critical ratios obtained when the means were compared for the Stone sample and the two Springfield samples. Since neither Brown nor Cantril and Allport reported standard deviations, it has not been possible to make similar comparisons utilizing their data.

It is customary to interpret the *Study of Values* in terms of the "average" score of 30<sup>2</sup> which must necessarily result. The Cantril-Allport and the Stone profiles are reported to provide a normative profile for men in general. All of Stone's cases and presumably the Cantril-Allport cases were college men. One must also remember that the scores represent relative interest rather than a measure of absolute interest along a given motivation axis.

<sup>2</sup> The manner of assigning weights to responses in the *Study of Values* is such that each respondent has a total of 180 points distributed among the six categories. The "average" of 30 is therefore a neutral score rather than the mean of any group with respect to a single category.

The H & PE profile appears about as expected. There is a peak in the *political* category; the Springfield men apparently have equally high *religious* motivation, with their third dominant interest in the social field. Brown's sampling, from a different college, shows peaks in the *political* and *social* areas with *religious* interest definitely subordinate.

The data in Table 2 show that the obtained differences in scores for the six categories represent real differences in interests as measured by the *Study of Values*. The critical ratios are large enough to indicate that most of these differences are not due to chance. It is unfortunate that direct comparisons cannot be made with the Cantril-Allport data, but it can be noted that the same general conclusions would be reached if one assumes their sample had similar SD's to those of Stone's sample.

One may summarize the data by stating that the *Study of Values* possesses considerable potentiality for guidance purposes. For example, one should expect H & PE applicants to be high in *political* and *social* motivation with the religious emphasis of the Springfield men being a matter of only local importance. Conversely, one would expect these men not to show peaks in both *political* and *economic* motivation; businessmen and lawyers are typically political-economic men (3, 4), while H & PE men are typically political-social or political-religious. One can also expect lower scores in *aesthetic* interest than are typical for men in general. The *theoretical* interest does not seem to be significant one way or another.

The Social Science profile is different not only from that of the H & PE men, but also from the pattern of college men in general. The Social Science peaks, as expected, are in the *social* and *religious* fields. The *political* and *economic* drives of the men seem to be relatively subordinate in comparison with those of the Cantril-Allport and Stone men. None of Stone's subsamples (Business, Banking, Medicine, Education, Law, or Literature) approached the Social Science profile in shape.

It would appear that the *Study of Values* can be useful in vocational counseling with men who are considering the Applied Social Sciences. One would stress relatively high *social* and *religious* motivation and relatively low *political* and *economic*

motivation. In common with most men, their *aesthetic* interests would be quite low while their *theoretical* interests might be average and not differential with respect to other men. H & PE and Social Science men differ primarily with respect to the inverse relation of the *political* variable. While both groups are definitely *religious* and probably more *social* than college men in general, they differ from each other greatly in *political* interest.

*Clinical Reports to Show the Usefulness of the Study of Values in Educational-Vocational Guidance*

The profiles of eleven men are noted in Table 3. Brief resumes of their careers in and after college are given.

TABLE 3  
*Profiles of Eleven College Students*

Case	Theoretical	Economic	Aesthetic	Social	Political	Religious
FC	32	23	20	36	18	51
AP	29	34	27	25	39	26
JP	33	20	17	34	38	38
AS	23	30	37	19	31	40
CD	42	27	28	35	22	26
PE	42	27	28	35	22	26
DS	25	27	34	31	24	39
RS	24	22	24	38	36	36
IW	44	26	25	32	17	36
AB	23	41	32	26	31	27
JG	32	16	27	42	18	45

- FC.* Brilliant student, who thinks of H & PE in terms of biological science. Good athlete in a few sports. Transferred to Social Sciences and is now a minister, which is consistent with his sophomore profile on the *Study of Values*.
- AP.* Did not return after one year—finances. Older man; found difficulty in adjusting to fellows. Considered “cussed type.” Tried all kinds of ways to return without money. Note strong economic and power drive and low social and religious motivation.
- JP.* Top-flight scholar and excellent athlete. Colored. Advised him to plan for graduate work in scientific aspects of H & PE. Considered ideal, all-around man on campus. Student assistant in a science laboratory.
- AS.* Flunked in another college, and wanted to “make a try” at social work. Musical. Family pressure to achieve.

- No follow-up data. There was considerable question about his choices, apart from his academic difficulties.
- CD.* Small, quiet, bright student who earned some letters in sports. Took master's degree in H & PE. Urged to continue for doctorate in tests and measurements aspects of H & PE. Teaching high school up to war. Note low political drive; he was not interested much in coaching but preferred teaching activity classes and health courses.
- PE.* Almost a straight-A student, but always an adjustment problem due to inability to crystallize plans. Came for social work, shifted to pre-medicine, and ended up planning graduate work in History. Consistent with *theoretical* and *social* scores.
- DS.* Good student in H & PE but not an athlete. In campus activities he was more like social science men. No record of activities along aesthetic lines, except membership in the Glee Club.
- RS.* B-plus student in spite of alleged disinterest in intellectual matters; genuinely an activity man. Profile good for H & PE at high level of accomplishment.
- IW.* H & PE man only because he had started in that field and did not want to lose credits, and did want to do recreational and YMCA physical work. Took as many Applied Social Science courses as he could get, asking for exceptions from prescribed curriculum. Did not earn letter and as far as the author knows did not try out for sports; yet considered for an assistantship as activity teacher.
- AB.* A definite misfit—truly an "economic man." Got jobs in student store and got into trouble on his jobs by being too managerial. Transferred to another college for business training; also because it was a "cheaper school."
- JG.* Social Science major, came up the hard way from wrong side of tracks through boys clubs and Y's. Good student and hard worker. Graduated with distinction and always had his eye on getting ahead in social field. Managed a team.

### Summary

1. The *Study of Values* has been shown to yield the expected profiles for groups of Health and Physical Education majors and Social Science majors in a men's undergraduate professional school.
2. The average Health and Physical Education profile shows high *political* motivation combined with either or both



*social* and *religious* motivation, while it is relatively low in comparison with the profile of men in general for *economic* and *aesthetic* motivation.

3. The Social Science profile is high in *social* and *religious* motivation and is relatively low in *political*, *economic*, and *aesthetic* motivation in comparison with the profile of men in general.

4. The "average" of 30 for the six categories, which is required by the nature of the questionnaire itself, cannot be considered a norm since it is not based upon averages scores of groups of people on each of the interest areas. Also, because of the sex differences reported by the test authors, it is important to compare men's profiles with the generalized profiles secured from men only.

5. With proper caution, the *Study of Values* is usable in vocational counseling and, because of its brevity and self-scoring features, it is a valuable adjunct to other methods of appraising interests through questionnaires and interviewing.

#### REFERENCES

1. Allport, G. W. and Vernon, P. E. *Manual for A Study of Values*. Boston: Houghton Mifflin Co., 1931.
2. Brown, M. "A Study of Personality of College Students of Physical Education." *Research Quarterly*, VI (1935), 69-77.
3. Cantril, H. and Allport, G. W. "Recent Applications of the Study of Values." *Journal of Abnormal and Social Psychology*, XXVIII (1933), 259-273.
4. Stone, C. L. "The Personality Factor in Vocational Guidance." *Journal of Abnormal and Social Psychology*, XXVIII (1933), 274-275.
5. Vernon, P. E. and Allport, G. W. "A Test for Personal Values." *Journal of Abnormal and Social Psychology*, XXVI (1931), 231-248.



# ENTRANCE EXAMINATIONS AT THE CITY COLLEGE OF NEW YORK

LOUIS LONG and JAMES D. PERRY  
City College of New York

## I. *Introduction*

THE use of psychological tests as entrance examinations at City College began in February, 1929, when the Student Personnel Bureau (now the Division of Testing and Guidance of the Department of Student Life) was called upon to administer tests to students whose high-school averages fell below the admission mark set by the College's Committee on Admission. The specific test batteries used from one year to another need not be described, but some of the tests used will be listed: *A.C.E. Psychological Examination*, *Army Alpha*, *College Entrance Examination*, *Thorndike Intelligence Test*, *Cooperative General Achievement Tests*, *Cooperative English Test*, and *Cooperative Contemporary Affairs Test*. In 1943 the College began to administer tests developed by K. W. Vaughn of the Carnegie Foundation and of the Graduate Record Office. The present battery of entrance examinations consists largely of tests constructed by Vaughn and included in his *Inventory of Scholastic Ability* (referred to hereafter as the Carnegie Inventory). The following tests are taken by all entering students: A.C.E. and four tests of the Carnegie Inventory (I—*General Verbal Ability*, II—*Technical Verbal Ability*, IV or IV-X—*General Mathematical Ability*, and VIII—*Social Studies Vocabulary*). In addition the science and engineering students take Test III—*Ability to Comprehend Scientific Materials* and all other students take Test IX—*Comprehension of Social Studies Materials*; both of these tests are also part of the Inventory. Finally, the engineering students take Test VI—*Spatial Visualizing Ability*.<sup>1</sup>

---

<sup>1</sup> Tests I, II, III, IV, and VI are from the *Pra-Engineering Inventory*.

At present all students are admitted to the College on the basis of a Composite Score, which includes both high-school average and test scores.

## II. *Evaluating the Effectiveness of the Entrance Examinations*

Determining the effectiveness of the Entrance Examination is a continuous process. The changes in the battery of tests used have been made on the basis of statistical studies carried on by members of the Division of Testing and Guidance.

One of the earliest studies undertaken was concerned with the effectiveness of a composite score computed from the scores on: (1) the *College Entrance Examination*, and (2) subject-matter examinations in English, a foreign language, and Mathematics. Correlations between the average of all grades in college and scores on the different parts of the Entrance Exam-

TABLE 1  
*Correlations between Average of All College Grades and Various  
Components of the Composite Score  
(N = 349)*

---

Average of all college grades vs. score on College Entrance Examination,	$r = 0.18$
Average of all college grades vs. Mathematics Examination,	$r = 0.19$
Average of all college grades vs. English Examination,	$r = 0.07$
Average of all college grades vs. Composite Score,	$r = 0.28$

---

inations were calculated for 349 students who were admitted to the College between September, 1932, and September, 1936, on the basis of their Composite Score on the Entrance Examination. The correlations reported by James D. Perry and Arthur L. Benton ranged from 0.07 to 0.28 (see Table 1).

When it appeared likely that the rental fee for the *College Entrance Examination* was going to be increased, it was decided to try other tests. Consequently, in February, 1939, the *Thorndike Intelligence Examination* was introduced but the scores on the examination supplied very little information about the student's ability that could not have been obtained from the results of the A.C.E. which all students took upon entering the college. Consequently, the Thorndike Examination was discontinued. It was felt that the Entrance Examination

should furnish more specific information about the student's abilities in various areas of knowledge, e.g., Social Science, Natural Science, Mathematics, English. With this in mind the Cooperative Tests were introduced in September, 1941. As data were collected, however, it became evident that some of the tests correlated so highly with each other<sup>2</sup> that including them all added very little to the effectiveness of the battery. Consequently, the *Cooperative English Test* was discontinued in February, 1943. Other studies indicated that the Cooperative Tests being used were not, in general, difficult enough for our students and that they did not differentiate sufficiently between the good and the poor students. Consequently, other agencies which might be in a position to supply the College with more appropriate tests were contacted. At about this time K. W. Vaughn was starting on a project, sponsored jointly by the Carnegie Foundation and two engineering societies, to construct entrance examinations appropriate for colleges of engineering; the battery is now known as the *Pre-Engineering Inventory* and some of the tests are included in the *Inventory of Scholastic Ability*. The College was able to try some of the Carnegie Tests on an experimental basis in February, 1943. The tests were revised on the basis of data collected and the new tests were administered to entering students in June and September, 1943. At this point these tests were no longer in an experimental state, but there had not yet been time to study their effectiveness in predicting success in college. Consequently, the results were not used in determining whether or not a student was to be admitted to the College.

Since both the Cooperative Tests and the Carnegie Tests were administered to the entering class of September, 1943, an excellent opportunity to compare the effectiveness of the various tests presented itself. First-term grades were collected as soon as they were available and an exhaustive correlation study was made. A summary of the correlations is presented in Table 2. It will be seen that the best single measure to use in predicting first-term grades is the high-school average; the

---

<sup>2</sup> The correlation between the total score on the *Cooperative English* and the L Score of the A.C.E. was 0.72.

one exception to this is the case of the B.S. students where test scores predict college success a bit better than does the high-school average. It is to be noted that in all three categories of students in Table 2 one of the Carnegie Tests will predict first-term grades more accurately than any of the Cooperative Tests. The best test to use with engineering students is the *Carnegie Comprehension of Scientific Materials*; with the B.S. students it is the *Carnegie Technical Vocabulary*; and with the entire freshman group it is the *Carnegie Comprehension*

TABLE 2

*Relationship between Test Scores, High-School Averages, and First-Term Averages for Students Entering the College in September, 1943*

Variables	Correlation Coefficients* First-Term Averages		
	Engineering Students ( <i>N</i> = 385)	B.S. Students ( <i>N</i> = 135)	All Students ( <i>N</i> = 662)
High-School Average .....	.52	.44	.54
A.C.E. Psychological Examination .....	.32	.50	.29
Cooperative C.A.			
I—Social Studies .....	.22	.34	.32
II—Natural Sciences .....	.32	.40	.34
III—Mathematics .....	.43	.40	.43
Carnegie Inventory			
I—General Verbal Ability .....	.33	.47	.37
II—Technical Verbal Ability .....	.37	.48	.35
III—Comprehension of Scientific Materials ..	.50	.40	.49

\* Product moment correlations are reported for engineering students; all others are tetrachoric correlations.

of *Scientific Materials*. The use of both high-school grades and test scores will, of course, increase the accuracy of prediction. The correlation between the Composite Scores and the first-term averages was found to be 0.57 for all students. This Composite Score included the high-school average and four separate test scores. It was found that just as accurate a prediction of the first-term grades for engineering students could be made by using the scores on only one test (the *Carnegie Comprehension of Scientific Materials*) and the high-school averages; this was almost true for the B.S. students (the correlations differed by 0.02). On the basis of the findings the

staff members of the Division recommended that in February, 1945, the Cooperative Tests be discontinued and that the results of the following Carnegie Tests be used in calculating the Composite Score: Test II—*Technical Verbal Ability*; Test IV—X—*General Mathematical Ability*; Test VIII—*Social Studies Vocabulary*. The scores on the A.C.E. and the high school averages were also included in the Composite Score, since it has always been the policy to admit students on the basis of a common battery. The philosophy behind this policy

TABLE 3

*Relationship between Test Scores, High-School Averages, and First-Term Averages for Students Entering the College in February, 1945*

Variables	Tetrachoric Correlation Coefficients First-Term Averages			
	Engi- neering Students ( <i>N</i> = 259)	B.S. Students ( <i>N</i> = 130)	B.B.A. Students ( <i>N</i> = 124)	B.A., B.S.S., B.S. in Ed Students ( <i>N</i> = 121)
High-School Average .....	.54	.49	.64	.64
A.C.E. Psychological Examination ...	.34	.30	.36	.33
Carnegie Inventory				
II—Technical Verbal Ability ....	.56	.29	.46	.22
III—Comprehension of Scientific Materials .....	.47	.17	..	..
IV—X—General Mathematical Ability	.42	.42	.51	.34
VIII—Social Studies Vocabulary ...	.30	.21	.51	.36
IX—Comprehension of Social Stud- ies Materials .....	..	..	.44	.41
Composite Score .....	.70	.44	.79	.67

is being seriously questioned since it means that one of the tests found most effective in predicting engineering grades is not being used in selecting entering students.

When first-term grades were available for students admitted on the basis of the new Composite Score the correlation between first-term averages and the Composite Scores was computed; it was found to be 0.54 (*N* = 784). Other correlations for this battery of tests will be found in Tables 3 and 4. It should be noted that for students taking certain courses the correlation coefficient between the Composite Scores and the first term averages was as high as 0.79.

### III. Limiting Factors in Predicting College Grades

At City College applications for admission as matriculated students are accepted from only those students with a high-school average of 75 or above. Furthermore, it is usual for the Entrance Examinations to be taken by about twice as many students as can be accepted. Both of these conditions restrict the range of the test scores and the high-school averages that will be used in determining the effectiveness of the tests. In this report no correction for a restriction of the range has been applied to the correlations reported.<sup>3</sup>

TABLE 4

*Correlations between Test Scores and Grades in Mathematics and Physics for Students Entering the College in February, 1945*

Tests	Grades in Physics 7 ( <i>N</i> = 193)	Grades in Math. 7 ( <i>N</i> = 193)
II—Technical Verbal Ability .....	.49	..
III—Comprehension of Scientific Materials ...	.46	..
IV-X—General Mathematical Ability .....	..	.44

Another factor that causes trouble in predicting grades is the unreliability of grades themselves. For example, it has been found that the first-term averages correlate about 0.6 with the second-term averages. This will naturally vary from term to term and from one curriculum to another. However, it is not possible for the correlations between first-term averages and a combination of test scores and high-school averages to be higher than the correlations between grades themselves. Consequently, the correlations reported in the tables above must be evaluated in terms of this ceiling.

### IV. The Use of the Results of the Entrance Examinations for Guidance Purposes

In selecting tests to be included in the battery of Entrance

<sup>3</sup> A number of schools have consistently reported a higher predictive correlation due to a difference in their entrance requirements. More specifically, the range of the City College group is about three-fourths of the range found in some other schools. This in turn reduces the City College predictive correlation by 20 to 30 per cent; for example, the correlation between grades in Physics 7 and Test III, *Comprehension of Scientific Materials*, is reported as 0.46 in Table 4. This correlation corrected for range restriction becomes 0.57.



Examinations there has always been an attempt made to select tests that would not only select the most capable students, but that would also furnish information about the students' aptitudes and abilities. The strength and weakness of the individual student in several different areas are measured very well by the present battery of examinations. For example, the student aspiring to become an engineer is tested for all-around college ability (*A.C.E. Psychological Examination* and Test I—*General Verbal Ability*) as well as in mathematics (Test IV—*General Mathematical Ability*) and science, especially physics (Test II—*Technical Verbal Ability* and Test III—*Ability to Comprehend Scientific Materials*). In addition he is given a social science test (Test VIII—*Social Science Vocabulary*) so that contrasting information about his ability to work in this field will be available to the guidance division. If such a student ever inquires about his aptitude or ability to do good work in the School of Technology, for example, considerable information on which to base an answer is already available.

In general terms the present battery of tests gives an overall appraisal for work in engineering, science, or social science (including business administration). Naturally in any counseling situation test scores are only one of several factors that are considered in giving a student an appraisal of his ability to succeed in a particular field.

The discussion in this paper has been limited to a consideration of the rôle that entrance examinations play in determining which students shall be admitted to the College. No attempt has been made to give a complete picture of how the results of the examinations are used in individual counseling situations. However, much of the statistical work is done with the idea of supplying the guidance personnel with more information about the effectiveness of a particular test and with other data that will help in counseling a particular student. The relationship between tests and courses is determined (see Table 4). Tables have been set up to show the effect of previous training on the test scores so that corrections may be made in evaluating the aptitudes and abilities of students with different educational backgrounds. Probability tables showing the chances

of an entering student with a particular Composite Score making a certain first-term average have been developed. A profile sheet, giving the score on each test and the percentile equivalent in terms of entering students, is prepared for each freshman. To recapitulate, the tests included in the Entrance Examinations are expected to serve a two-fold purpose: (1) to select the most capable students, and (2) to determine the aptitudes and abilities of the individual student.

# SELECTING EXECUTIVES BY PSYCHOLOGICAL TESTS

CLAUDE EDWARD THOMPSON

University of Omaha

THE purpose of the research reported here was to determine what contribution psychological tests could make to the personnel procurement procedures of one of the country's largest firms of Consulting Management Engineers.

## *The Criterion*

Interviews were held with Partners, Account Managers, Department Managers, Account Supervisors, and with Staff Men. These interviews had the following aims:

1. To get the consensus concerning the demands made on staff men by the work of the firm.
2. To define the abilities, personal characteristics, and accomplishments considered necessary for success with the firm.

Two groups of men were selected as the experimental groups to be analyzed. These men were selected as follows:

1. Those men whose performance records and unanimous partner-evaluations placed them as "superior" constituted one group. There were fifteen men in this group.
2. Those men whose performance records and unanimous partner-evaluations placed them at or below average were called the "average" men. There were ten men in this group.

Therefore, the criterion consisted of performance records and consensus of partner-evaluations combined into a single standing expressed as "superior" or "average." The criterion so obtained is a two-category variable which does not arrange the men in rank order. Therefore it does not readily lend itself to the statistical technique of correlation. However, group comparisons can be made which will show whether or not there

- No follow up data. There was considerable question about his choice, apart from his academic difficulties.
- CD.* Small, quiet, bright student who earned some letters in sports. Took master's degree in H & PE. Urged to continue for degreeate in tests and measurements aspects of H & PE. Teaching high school up to now. Note low quality of work, he was not interested much in teaching but preferred teaching activity classes and health courses.
- PE.* Almost a straight-A student, but always an adjustment problem due to inability to crystallize plans. Came for social work, shifted to pre medicine, and ended up planning graduate work in History. Consistent with theoretical and social scores.
- DS.* Good student in H & PE but not an athlete. In campus activities he was more like social science men. No record of activities along athletic lines, except membership in the Yale Club.
- RS.* B-plus student in spite of alleged distastes in intellectual matters, genuinely an activity man. Profile good for H & PE at high level of accomplishment.
- HW.* H & PE man only because he had started in that field and did not want to lose credits, and did want to do recreational and YMCA physical work. Took as many Applied Social Science courses as he could get, asking for exceptions from premedical curriculum. Did not earn letters and as far as the author knows did not try out for sports, yet completed for an assistantship as activity teacher.
- AB.* A definite misfit, truly an "economic man." Got jobs in student store and got into trouble on his job by being too managerial. Transferred to another college for business training, also because it was a "cheaper school."
- HC.* Social Science major, came up the hard way from wrong side of tracks through boys clubs and Y's. Good student and hard worker. Graduated with distinction and always had his eye on getting ahead in social field. Managed a team.

#### Summary

1. The Study of Values has been shown to yield the expected profiles for groups of Health and Physical Education majors and Social Science majors in a men's undergraduate professional school.
2. The average Health and Physical Education profile shows high *political* motivation combined with either or both

*social* and *religious* motivation, while it is relatively low in comparison with the profile of men in general for *economic* and *aesthetic* motivation.

3. The Social Science profile is high in *social* and *religious* motivation and is relatively low in *political*, *economic*, and *aesthetic* motivation in comparison with the profile of men in general.

4. The "average" of 30 for the six categories, which is required by the nature of the questionnaire itself, cannot be considered a norm since it is not based upon averages scores of groups of people on each of the interest areas. Also, because of the sex differences reported by the test authors, it is important to compare men's profiles with the generalized profiles secured from men only.

5. With proper caution, the *Study of Values* is usable in vocational counseling and, because of its brevity and self-scoring features, it is a valuable adjunct to other methods of appraising interests through questionnaires and interviewing.

#### REFERENCES

1. Allport, G. W. and Vernon, P. E. *Manual for A Study of Values*. Boston: Houghton Mifflin Co., 1931.
2. Brown, M. "A Study of Personality of College Students of Physical Education." *Research Quarterly*, VI (1935), 69-77.
3. Cantril, H. and Allport, G. W. "Recent Applications of the Study of Values." *Journal of Abnormal and Social Psychology*, XXVIII (1933), 259-273.
4. Stone, C. L. "The Personality Factor in Vocational Guidance." *Journal of Abnormal and Social Psychology*, XXVIII (1933), 274-275.
5. Vernon, P. E. and Allport, G. W. "A Test for Personal Values." *Journal of Abnormal and Social Psychology*, XXVI (1931), 231-248.



# ENTRANCE EXAMINATIONS AT THE CITY COLLEGE OF NEW YORK

LOUIS LONG and JAMES D. PERRY  
City College of New York

## I. Introduction

THE use of psychological tests as entrance examinations at City College began in February, 1929, when the Student Personnel Bureau (now the Division of Testing and Guidance of the Department of Student Life) was called upon to administer tests to students whose high-school averages fell below the admission mark set by the College's Committee on Admission. The specific test batteries used from one year to another need not be described, but some of the tests used will be listed: *A.C.E. Psychological Examination*, *Army Alpha*, *College Entrance Examination*, *Thorndike Intelligence Test*, *Cooperative General Achievement Tests*, *Cooperative English Test*, and *Cooperative Contemporary Affairs Test*. In 1943 the College began to administer tests developed by K. W. Vaughn of the Carnegie Foundation and of the Graduate Record Office. The present battery of entrance examinations consists largely of tests constructed by Vaughn and included in his *Inventory of Scholastic Ability* (referred to hereafter as the Carnegie Inventory). The following tests are taken by all entering students: A.C.E. and four tests of the Carnegie Inventory (I—*General Verbal Ability*, II—*Technical Verbal Ability*, IV or IV-X—*General Mathematical Ability*, and VIII—*Social Studies Vocabulary*). In addition the science and engineering students take Test III—*Ability to Comprehend Scientific Materials* and all other students take Test IX—*Comprehension of Social Studies Materials*; both of these tests are also part of the Inventory. Finally, the engineering students take Test VI—*Spatial Visualizing Ability*.<sup>1</sup>

---

<sup>1</sup> Tests I, II, III, IV, and VI are from the *Pre-Engineering Inventory*

At present all students are admitted to the College on the basis of a Composite Score, which includes both high-school average and test scores.

## II. *Evaluating the Effectiveness of the Entrance Examinations*

Determining the effectiveness of the Entrance Examination is a continuous process. The changes in the battery of tests used have been made on the basis of statistical studies carried on by members of the Division of Testing and Guidance.

One of the earliest studies undertaken was concerned with the effectiveness of a composite score computed from the scores on: (1) the *College Entrance Examination*, and (2) subject-matter examinations in English, a foreign language, and Mathematics. Correlations between the average of all grades in college and scores on the different parts of the Entrance Exam-

TABLE 1

*Correlations between Average of the College Grades and Entrance Examination Scores, 1932-1936*

Average of all college grades vs. score on College Entrance Examination	$r = 0.18$
Average of all college grades vs. Mathematics Examination	$r = 0.19$
Average of all college grades vs. English Examination	$r = 0.07$
Average of all college grades vs. Composite Score	$r = 0.28$

inations were calculated for 349 students who were admitted to the College between September, 1932, and September, 1936, on the basis of their Composite Score on the Entrance Examination. The correlations reported by James D. Perry and Arthur L. Benton ranged from 0.07 to 0.28 (see Table 1).

When it appeared likely that the rental fee for the *College Entrance Examination* was going to be increased, it was decided to try other tests. Consequently, in February, 1939, the *Thorndike Intelligence Examination* was introduced but the scores on the examination supplied very little information about the student's ability that could not have been obtained from the results of the A.C.E., which all students took upon entering the college. Consequently, the Thorndike Examination was discontinued. It was felt that the Entrance Examination



should furnish more specific information about the student's abilities in various areas of knowledge, e.g., Social Science, Natural Science, Mathematics, English. With this in mind the Cooperative Tests were introduced in September, 1941. As data were collected, however, it became evident that some of the tests correlated so highly with each other<sup>2</sup> that including them all added very little to the effectiveness of the battery. Consequently, the *Cooperative English Test* was discontinued in February, 1943. Other studies indicated that the Cooperative Tests being used were not, in general, difficult enough for our students and that they did not differentiate sufficiently between the good and the poor students. Consequently, other agencies which might be in a position to supply the College with more appropriate tests were contacted. At about this time K. W. Vaughn was starting on a project, sponsored jointly by the Carnegie Foundation and two engineering societies, to construct entrance examinations appropriate for colleges of engineering; the battery is now known as the *Pre-Engineering Inventory* and some of the tests are included in the *Inventory of Scholastic Ability*. The College was able to try some of the Carnegie Tests on an experimental basis in February, 1943. The tests were revised on the basis of data collected and the new tests were administered to entering students in June and September, 1943. At this point these tests were no longer in an experimental state, but there had not yet been time to study their effectiveness in predicting success in college. Consequently, the results were not used in determining whether or not a student was to be admitted to the College.

Since both the Cooperative Tests and the Carnegie Tests were administered to the entering class of September, 1943, an excellent opportunity to compare the effectiveness of the various tests presented itself. First-term grades were collected as soon as they were available and an exhaustive correlation study was made. A summary of the correlations is presented in Table 2. It will be seen that the best single measure to use in predicting first-term grades is the high-school average; the

---

<sup>2</sup> The correlation between the total score on the Cooperative English and the L Score of the A.C.E. was 0.72.

one exception to this is the case of the B.S. students where test scores predict college success a bit better than does the high school average. It is to be noted that in all three categories of students in Table 2 one of the Carnegie Tests will predict first term grades more accurately than any of the Cooperative Tests. The best test to use with engineering students is the *Carnegie Comprehension of Scientific Materials*; with the B.S. students it is the *Carnegie Technical Vocabulary*; and with the entire freshman group it is the *Carnegie Comprehension*.

TABLE 2

Predicting Performance in the High School Grades and First Term Averages for Students Entering the Engineering School

Tests	Correlation Coefficients*		
	First Term Averages		
	Engineering Students (N = 111)	B.S. Students (N = 111)	All Students (N = 222)
High School Average	.42	.44	.54
C. C. Psychological Examination	.42	.49	.29
C. C. Composite Score			
I - High School Average	.42	.44	.54
II - Technical Vocabulary	.42	.49	.34
III - Mathematics	.42	.40	.33
Carnegie Examinations			
I - General Mental Ability	.42	.47	.37
II - Technical Vocabulary	.42	.48	.35
III - Comprehension of Scientific Materials	.44	.40	.49

\* Pearson's coefficient correlations are reported for engineering students; all others are statistical correlations.

of *Scientific Materials*. The use of both high-school grades and test scores will, of course, increase the accuracy of prediction. The correlation between the Composite Scores and the first-term averages was found to be 0.57 for all students. This Composite Score included the high-school average and four separate test scores. It was found that just as accurate a prediction of the first-term grades for engineering students could be made by using the scores on only one test (the *Carnegie Comprehension of Scientific Materials*) and the high-school averages; this was almost true for the B.S. students (the correlations differed by 0.02). On the basis of the findings the

staff members of the Division recommended that in February, 1945, the Cooperative Tests be discontinued and that the results of the following Carnegie Tests be used in calculating the Composite Score: Test II—*Technical Verbal Ability*; Test IV—X—*General Mathematical Ability*; Test VIII—*Social Studies Vocabulary*. The scores on the A.C.E. and the high school averages were also included in the Composite Score, since it has always been the policy to admit students on the basis of a common battery. The philosophy behind this policy

TABLE 3

*Relationship between Test Scores, High-School Averages, and First-Term Averages for Students Entering the College in February, 1945*

Variables	Tetrachoric Correlation Coefficients First-Term Averages			
	Engi- neering Students ( <i>N</i> = 259)	B.S. Students ( <i>N</i> = 130)	B.B.A. Students ( <i>N</i> = 124)	B.A., B.S.S., B.S. in Ed Students ( <i>N</i> = 121)
High-School Average . . . . .	.54	.49	.64	.64
A.C.E. Psychological Examination . . . .	.34	.30	.36	.33
Carnegie Inventory				
II—Technical Verbal Ability . . . .	.56	.29	.46	.22
III—Comprehension of Scientific				
Ability . . . . .	.47	.17		
I . . . . .	.42	.42	.51	.34
lary . . . . .	.30	.21	.51	.36
IX—Comprehension of Social Stud-				
ies Materials . . . . .	..	..	.44	.41
Composite Score . . . . .	.70	.44	.79	.67

is being seriously questioned since it means that one of the tests found most effective in predicting engineering grades is not being used in selecting entering students.

When first-term grades were available for students admitted on the basis of the new Composite Score the correlation between first-term averages and the Composite Scores was computed; it was found to be 0.54 (*N* = 784). Other correlations for this battery of tests will be found in Tables 3 and 4. It should be noted that for students taking certain courses the correlation coefficient between the Composite Scores and the first term averages was as high as 0.79.

### III. Limiting Factors in Predicting College Grades

At City College applications for admission as matriculated students are accepted from only those students with a high-school average of 75 or above. Furthermore, it is usual for the Entrance Examinations to be taken by about twice as many students as can be accepted. Both of these conditions restrict the range of the test scores and the high school averages that will be used in determining the effectiveness of the tests. In this report no correction for a restriction of the range has been applied to the correlations reported.<sup>1</sup>

TABLE 4

*Correlations between Test Scores and Grades in Mathematics and Physics for Seniors, Entering the College in February, 1945*

Tests	Grades in Physics 7 (N = 195)	Grades in Math 7 (N = 193)
II Technical Total Ability	.49	..
III Comprehensive Written Materials	.46	..
IV A General Mathematical Ability	..	.44

Another factor that causes trouble in predicting grades is the unreliability of grades themselves. For example, it has been found that the first-term averages correlate about 0.6 with the second term averages. This will naturally vary from term to term and from one curriculum to another. However, it is not possible for the correlations between first-term averages and a combination of test scores and high-school averages to be higher than the correlations between grades themselves. Consequently, the correlations reported in the tables above must be evaluated in terms of this ceiling.

### IV. The Use of the Results of the Entrance Examinations for Guidance Purposes

In selecting tests to be included in the battery of Entrance

<sup>1</sup> A number of schools have consistently reported a higher predictive correlation due to a difference in their entrance requirements. More specifically, the range of the City College group is about three-fourths of the range found in some other schools. This in turn reduces the City College predictive correlation by 20 to 30 per cent; for example, the correlation between grades in Physics 7 and Test III, *Comprehension of Scientific Materials*, is reported as 0.46 in Table 4. This correlation corrected for range restriction becomes 0.57.

Examinations there has always been an attempt made to select tests that would not only select the most capable students, but that would also furnish information about the students' aptitudes and abilities. The strength and weakness of the individual student in several different areas are measured very well by the present battery of examinations. For example, the student aspiring to become an engineer is tested for all-around college ability (*A.C.E. Psychological Examination* and Test I—*General Verbal Ability*) as well as in mathematics (Test IV—*General Mathematical Ability*) and science, especially physics (Test II—*Technical Verbal Ability* and Test III—*Ability to Comprehend Scientific Materials*). In addition he is given a social science test (Test VIII—*Social Science Vocabulary*) so that contrasting information about his ability to work in this field will be available to the guidance division. If such a student ever inquires about his aptitude or ability to do good work in the School of Technology, for example, considerable information on which to base an answer is already available.

In general terms the present battery of tests gives an overall appraisal for work in engineering, science, or social science (including business administration). Naturally in any counseling situation test scores are only one of several factors that are considered in giving a student an appraisal of his ability to succeed in a particular field.

The discussion in this paper has been limited to a consideration of the rôle that entrance examinations play in determining which students shall be admitted to the College. No attempt has been made to give a complete picture of how the results of the examinations are used in individual counseling situations. However, much of the statistical work is done with the idea of supplying the guidance personnel with more information about the effectiveness of a particular test and with other data that will help in counseling a particular student. The relationship between tests and courses is determined (see Table 4). Tables have been set up to show the effect of previous training on the test scores so that corrections may be made in evaluating the aptitudes and abilities of students with different educational backgrounds. Probability tables showing the chances

of an entrance student a "Composite Score" making a comparison to standards have been developed. A profile sheet, giving the score on each test and the percentile equivalent of scores of entering students, is prepared for each freshman. In summary, the tests included in the Entrance Examination are expected to serve a two-fold purpose: (1) to select the more capable students, and (2) to determine the aptitudes and abilities of the individual student.

# SELECTING EXECUTIVES BY PSYCHOLOGICAL TESTS

CLAUDE EDWARD THOMPSON

University of Omaha

THE purpose of the research reported here was to determine what contribution psychological tests could make to the personnel procurement procedures of one of the country's largest firms of Consulting Management Engineers.

## *The Criterion*

Interviews were held with Partners, Account Managers, Department Managers, Account Supervisors, and with Staff Men. These interviews had the following aims:

1. To get the consensus concerning the demands made on staff men by the work of the firm.

2. To define the abilities, personal characteristics, and accomplishments considered necessary for success with the firm.

Two groups of men were selected as the experimental groups to be analyzed. These men were selected as follows:

1. Those men whose performance records and unanimous partner-evaluations placed them as "superior" constituted one group. There were fifteen men in this group.

2. Those men whose performance records and unanimous partner-evaluations placed them at or below average were called the "average" men. There were ten men in this group.

Therefore, the criterion consisted of performance records and consensus of partner-evaluations combined into a single standing expressed as "superior" or "average." The criterion so obtained is a two-category variable which does not arrange the men in rank order. Therefore it does not readily lend itself to the statistical technique of correlation. However, group comparisons can be made which will show whether or not there





dall); *The Personal Audit* (Form LL) (C.R. Adams); *Preference Record* (Form BB) (G. Frederic Kuder); the *A-S Reaction Study-Revised for Business Use* (R. O. Beckman); *The Root I-E Test*; *The Guilford-Martin Personnel Inventory*.

### *Data Analysis and Results*

The average scores and standard deviations for the two groups, "superior" and "average," were computed from the obtained scores on all tests. The average scores and standard deviations were converted into percentile ranks. These data are presented in Table 1. Inspection of Table 1 shows the average standing on each test for superior men and average men, expressed on a 100-point scale. It also shows the scores equivalent to  $\pm 1 \sigma$  and  $-1 \sigma$  converted into percentile ranks. In order to determine whether or not there were any real differences between the average scores of the two groups of men on the tests, the ratio of the difference in mean score to the standard error of the difference was obtained for each test. This ratio determines the chances in 100 that an obtained difference between means is a true difference rather than due to chance operating factors.

In Table 2 are shown the "critical" ratios and the chances in 100 that each ratio represents a true difference in mean test score between superior men and average men. Due to the highly-selected character of the 25 men used in this research, it is probable that the group of men called "average" in intra-firm comparisons are "superior" by extra-firm requirements or by less rigid standards of excellence than those used. If these possibilities were true, significant differences between the two groups might have been of still greater magnitude had it been possible to evaluate the tests using a less homogeneous sample of men.

These considerations are presented because, "It is customary to take a  $D/\sigma_p$  of 3.00 as indicative of a significant difference (virtual certainty) since there is only about one chance in 1000 that a difference of  $\pm 3$  will arise when the true difference is zero." This assumes a wide range of scores in whatever trait is being measured. In this case it was rather arbitrarily

decided to consider a difference as probably true if there were at least 90 chances in 100 of a true difference. On this basis, the two groups of men differ significantly on the following tests:

	Chances in 100 of a true difference
<i>Wonderlich Personnel Test</i> .....	96
<i>Michigan V.P. Test (Total Score)</i> .....	98
Government .....	97
Physical Science .....	96
Mathematics .....	100
Sports .....	99
<i>Kuder Preference Record</i>	
Mechanical .....	96
Social Service .....	99
<i>Personal Audit</i>	
Firmness-Indecision .....	98
Frankness-Evasion .....	93
Stability-Instability .....	100
Tolerance-Intolerance .....	91

The cases of several other measures require further analysis (See Tables 1 and 2). One of these is the Persuasive component of the *Kuder Preference Record*. Here the mean percentile standings and standard deviations are practically the same. The means of both groups of men are 94 (Superior) and 93 (Average) percentile. Though no statistically significant difference between the two groups occurs, the fact that the 25 men average so far above 50 percentile in Persuasive interest seems to indicate that they were made homogeneous in this respect in the original selection. Therefore, it is thought that the Persuasive component of the *Kuder Preference Record* should be used to aid in selecting the executives of this firm, and that rather high standing in this component should be considered desirable.

Somewhat the same situation exists with respect to Commerce vocabulary on the *Michigan Vocabulary Profile Test*, Literary interest on the *Kuder Preference Record*, and Tranquillity-Irritability and Contentment-Worry on the *Personal Audit*.

With respect to Commerce vocabulary, Persuasive and Literary interests, and the personality traits of Tranquillity-



terest the superior men average 51 percentile, while the average men average 79 percentile. There are 99 chances in 100 that this difference in favor of the average men is real.

Suppose the average percentiles for the superior group were used as the standard to which applicants' test scores were compared. The lower standard deviation limits could be used as the zone for acceptance. Scores falling much below this zone would then be focal points for further analysis of the applicant. This would be particularly true for those tests in which the superior men are very high and or in which the superior men are significantly superior to the average men.

### *Summary*

1. Eight separate psychological tests, comprising 34 different measures, were administered to 25 staff men of a Management Engineering firm.

2. Significant differences in mean scores were found on 12 of the 34 measures. It was also found that in five of the remaining 22 measures both groups were definitely above average, though the groups did not differ significantly between themselves.

3. The Superior Man is above average in 16 of the 17 measures and average in one. The Average Man is above average in 13 of the 17 measures, average in three, and below average in one. The Superior man has higher percentile standings than the Average Man on 13 of the 17 measures.

## A MISUSE OF GROUP TESTS OF INTELLIGENCE IN THE SCHOOL

D. A. WORCESTER  
The University of Nebraska

THE writer was asked a short time ago to visit a small school system in the state of Nebraska and to help analyze some of the reading problems which were occurring in that school.

The first case that was sent to him was a boy of 12 years and 9 months of age, classed in the 5th grade, who had been given a *Henmon-Nelson Test of Mental Ability* and had obtained an IQ rating on this test of 53. The assumption was that he was having difficulty with his reading because of his low mental ability and that he had been passed along to keep him nearer his chronological age group. He had also been given the *Iowa Silent Reading Test*, Elementary Test, Revised Form AM. This indicated a total grade equivalent in reading of 3.5, but there were some rather surprising scores on certain parts of the test. On Test 1, Comprehension, for example, his score indicated a grade level of 5.4 and on Test 4, Paragraph Comprehension, he obtained a grade level of 6.7, a rather high achievement for one with an IQ of 53.

It seemed wise to give this boy an individual mental test and on the *Stanford-Binet Test*, Form L, his IQ was 78. It was noticed while giving the test that the boy's language ability was about that of a 10-year-old boy. As a further check, then, the *Pintner-Paterson Scales*, Short Form, were given and upon these scales the boy attained a PQ of 98. Here then was an individual who on three intelligence tests given within a few months of each other had ratings of 53, 78, and 98, a variation of 45 points. Of course these IQ's and PQ's are not equivalent but these differences are far greater than mere normal variations between tests. The next step was to ask the boy to read aloud

from those portions of the *Low Silent Reading Test* upon which he had made scores equal to or better than his grade placement of 5. It became immediately apparent that he could hardly read at all. He stumbled over the simplest words and was exceedingly slow in reading the few words which he did know. He then was asked to read aloud some of the questions which he had answered correctly. This he was almost wholly unable to do. Furthermore, he had answered, and in some instances correctly answered, questions which were based on material in paragraphs far beyond the point to which he had read during the original test.

Additional questioning indicated that this lad had followed instructions to put marks in certain spaces to indicate his choices of correct answers to questions and by pure chance had attained scores which had given him a much higher rating on certain of the tests than his true abilities warranted.

It is clear that his very low score on the *Henmon-Nelson Test of Mental Ability* was a result of his inability to read rather than because of his low intelligence. Obviously it is not sure yet which of the 3 mental ability scores was the nearest correct in his case, but it seems likely that the true score lay somewhere between the IQ, 78, on the Stanford-Binet and the PQ, 98, on the Pintner-Paterson. An examination of the boy's social background and general experience revealed some good reasons for his poverty in language.

The second person who was referred for examination was a boy of 10 years and 2 months of age; also a non-reader and also one who had already been given a Henmon-Nelson test. His IQ on that test was 86. He was a boy with a slight speech defect, and who had had poliomyelitis 2 or 3 years earlier. On the Stanford-Binet, Form I, this fellow showed an IQ of 115. This is another startling example of the inadequacy of the typical group test for a person with a serious reading defect. In this particular instance the boy had an excellent speaking vocabulary. His Stanford-Binet vocabulary level was 2 years beyond his chronological age and in his conversation he exhibited an unusual fluency in language. However, he had not learned to recognize words which he knew.

The above experiences led the writer to note if similar test results were encountered elsewhere. Soon another case of serious reading disability was referred to him—a child who had been given the Kuhlmann-Anderson Test and had secured an IQ rating on it of 84. On the Stanford-Binet she had an IQ of 106. Still more recently, from a town of 50,000 with an alert modern school system, there came two individuals—one whose school record blank showed a *Pintner General Ability Test* IQ of 64 and one a Henmon-Nelson IQ of 62. Both were referred as reading problems. The first of these on Stanford-Binet, Form M, secured an IQ of 104, a difference of forty points; the second attained a Stanford-Binet, Form L, IQ of 90, a difference of 28 points. On each of these scores the Stanford-Binet rating was more likely to be truly representative of the real ability of the individual than was that of the group test.

Perhaps there is nothing new or startling to psychologists or to those who are engaged in remedial reading in these results. But we do have here illustrations of a very serious and apparently not uncommon practice in schools. Group tests of mental ability are widely used and in very many schools are given to all pupils. The results of these tests are referred to as an aid in solving educational problems. If a child is not learning well and the intelligence test shows him to be of low mental ability, the tendency is to assume that the answer is known and that not much help can be given.

It is apparent that many school people have not been trained to the point of realizing that the choice of the mental ability test appropriate for a given individual must be based upon a knowledge of the circumstances surrounding his case. The absurdity of giving a mental test involving reading to one who is deficient in reading and the great harm which is likely to result from such a practice should be obvious, but it seems that the caution must be given again and again.

Incidentally, the first case cited above shows that even on a group test of reading, it is possible to make a score widely at variance from the facts. It should be stated categorically, perhaps, that *no group test of any kind* should be used unless there be provision for intensive, individual study of those persons making low scores on it.





# MULTIPLE FORMS OF OBJECTIVE TESTS: A TEST-SCORING MACHINE TECHNIQUE

VERNON W. LEMMON and PHILIP H. DuBOIS  
Washington University

It is often desirable to use different forms of classroom subject-matter tests in order to reduce opportunities for collusion among students. A method which is now used routinely in General Psychology classes at Washington University permits the production of a number of forms of identical content from a single set of mimeograph stencils.

Essentially the method consists of assigning as many numbers to each question as there are forms. These numbers refer to the positions on the answer sheet where the answers are to be indicated rather than the order in which the questions appear on the duplicated material. All forms have the same items in the same order, although the numbering of the items is not consecutive but in an irregular order which is different for each form.

Student acceptance of the device of indicating answers in irregular order on the answer sheet has been good. The extra amount of time required is not great. Students have remarked that there is less opportunity of losing their places and of answering a number of consecutive questions in the wrong spaces since for each question attention must be given to finding the appropriate position on the answer sheet for indicating the answer.

The chief advantages of the method are three-fold: (1) reduction of the cost of duplicating forms identical in content; (2) insuring that the test material (including any typographical errors!) is identical for all subjects; (3) the answer sheets of students sitting next to each other show entirely different marking patterns, thus making copying difficult.

The following procedure has been used when the test items have already been typed on cards, with a letter indicating the correct answer, such as T or F; or A, B, C, D or E (for multiple-choice items):

1. The cards are numbered consecutively (in pencil) from 1 through the final item. These are the answer-position numbers that will indicate the positions on the answer sheet used by students taking Form A of the test. At this time a scoring key for Form A is prepared by punching out the appropriate holes in a machine-stencil.

2. The cards are thoroughly shuffled and the procedure followed in Step 1 is repeated, thus producing another series of answer-position numbers for Form B. These numbers are placed on the card beside the answer-position numbers for Form A. A machine key for Form B is then punched out, following the second answer-position numbers.

3. If more forms are desired, the procedure is repeated until the required number of sets of numbers (and keys) has been obtained. For classroom tests under crowded conditions we have found that three forms are adequate to prevent collusion.

4. The cards are given a final shuffle to determine the order in which the items will appear on the examination. This extra shuffle is to insure that the numbers for the final form will not be in consecutive order. A mimeograph stencil is then made from the cards, each item being preceded by a set of three numbers (a three-form test is chosen for illustration) in a horizontal row to the left of the item. Thus when the stencil sheet is completed, all of the A-form numbers will be in the left column, the B's next, etc. There should be a slight separation between columns. Letters identifying the forms should be placed at the head of each column on each page.

A small section of a typical mimeograph stencil is shown below:

A	B	C	
12	63	29	A trichromat is usually totally color-blind.
25	14	27	Some after-images can be used to determine the complement of a given color.
56	22	33	In binocular color mixture, one color is presented to one eye and the other color to the other eye.

In running off copies for the A-form, the B and C columns of numbers should be masked off the stencil with strips of stencil backing, and similarly for the other forms, so that only one column of numbers will appear on each sheet.

It is desirable to use different colors of paper for the several forms, thus guarding against the possibility of exchange of forms by students. When the examination consists of several sheets, it is advisable to assemble the sheets in a different order for each form.

Since the numbers have been placed on the cards in pencil, they can be easily erased when it is desired to use the same cards for another test.

Other methods of constructing the columns of numbers may of course be employed. For example, if the items have been typed on a sheet (with T or F in front of each), the A column may be filled in by assigning the numbers 1, 2, 3, etc., to items by skipping up and down the column in random fashion. The B and C columns may be constructed in a similar manner, the order in each column being of course independent of the others. Machine keys may then be punched for each column.

If a file of standard machine keys is kept, the number columns can be made to agree with previously chosen keys. This is probably easier with the "sheet" method just described. If the number 36, for example, is marked T on the standard key, care should be taken to put this number opposite some T item.

If new machine keys are to be made to fit the number columns, the "card" method has the advantage that the numbers to be punched on the key are in consecutive order before the cards are shuffled to make the next form, which makes it easier to punch the keys.



## BOOK REVIEW

Dorothy C. Adkins and Associates. *Construction and Analysis of Achievement Tests*. Washington: Superintendent of Documents, 1947. \$x.xx

A psychological experimenter is under obligation to indicate any source of bias in his experiment, even though he is confident that it has exerted no effect upon his results. This obligation is the more binding upon a reviewer. It may be that a number of years of close professional association with the senior author and her associates as co-workers in public personnel research and test construction have predisposed the reviewer to a favorable opinion of this book. It is the reviewer's firm conviction, however, that the effect of this association has been to heighten his awareness of the need which is filled by the book and his appreciation of the limiting conditions which necessitated some of its apparent shortcomings. Be that as it may, the reader is warned.

*Construction and Analysis of Achievement Tests* is written primarily from the viewpoint of civil service and merit system testing. As stated in the preface,

The original purpose of this book was to serve as a basis for training personnel of the U. S. Civil Service Commission and of the U. S. War Department who were to be directly engaged in the preparation of written or performance tests of achievement for predicting job performance. . . . It is believed that it may be of interest to other agencies engaged in examining work, to State and local civil service agencies and to private industry. Universities may also find it of value in preparing students for examining work.

To these beliefs the reviewer heartily subscribes. Although directed toward achievement testing in the merit-system setting, with all of its examples drawn from that field, there is no area of psychological testing to which most of the principal contents of the book are not applicable even though the technical detail is sometimes greater than is usually applied.

The book brings together basic principles and procedural details of test construction and analysis from widely scattered sources—professional journals, agency manuals and the accumulated experience of workers in the field which is not elsewhere set down. Careful attention to these principles and techniques could vastly improve the quality of merit-system examining and, by increasing the com-

potence level of public servant, could proportionately increase the quality of public service.

The stated purpose of the book has sharply limited its scope, though not always disadvantageously. To quote for the last time, "The sole basis for selection has been practicality. Concepts used in, or directly pertinent to test development are included. Purely theoretical considerations found no place." Included are such considerations as the appropriate length of tests, the need for maintaining the confidential nature of test items, and the appearance of practicality in items and methods of filing items. Also treated are such topics as reliability, validity, and weighting, with frequent references to Thurstone's *Reliability and Validity of Tests* and to *Psychometrics*. The emphasis on a practical approach has resulted in some unfortunate omissions and distortions of emphasis. Thus nearly as much attention is devoted to methods of filing items as to the development of a criterion. The fallacy of using a split-halves or Kuder-Richardson reliability estimate for speeded tests is nowhere mentioned; neither are the problems of item analysis of the final items in a time-limit test. Although special care has been taken to define, in an excellent glossary, the technical terms of statistics and psychological testing, equally esoteric terms from the federal personnel vocabulary are used without explanation. Terms such as "CAF-2," "classification," and "grade" may present difficulties in their technical meanings to readers not versed in the federal personnel system.

Chapter I deals with planning a written test, defines the characteristics of a good test, treats of the necessary cooperation of subject-matter specialists and test technicians, carries through the process of analyzing the job for testing purposes and of translating the job analysis into an outline of areas to be tested, both for single jobs and for groups of related jobs. Test planning is approached from the sampling viewpoint. Problems of administration and scoring as they affect test planning and arrangement are also covered.

In Chapter II, *Constructing and Compiling Written Tests*, attention is concentrated on limited-response items and focussed further on multiple-choice items, although other types are described and evaluated. The procedures, underlying logic, rules-of-thumb, and canons of good practice in writing limited response items are presented and illustrated. Procedures and standards for the review of items are set forth. Methods are described for converting an assemblage of items into a complete test designed, insofar as possible, to meet its analyzed objectives under the conditions of its use.

A selective introduction to the statistical concepts most often used in psychological testing constitutes the third chapter. It extends through partial and multiple correlation and elementary sampling theory and covers briefly tetrachoric, biserial and point-biserial correlation as well. There can be little disagreement with most of the concepts included, although the practical applications of kurtosis

or the standard error of a standard deviation are likely to be infrequent. Not all of the topics covered will, of course, find use in all test construction agencies. Some readers will feel that the discussion of sampling should have referred to the logic of the null hypothesis and have at least acknowledged the existence of such sampling statistics as  $t$ .

The Analysis of Test Results, following the basic statistics, gives an elementary, but thoroughly sound introduction to the concepts of norms, standardization, item difficulties, reliabilities, validity for a specified purpose, item analysis, weighting and the practical applications of analysis toward the improvement of the test. The presentation here, as in the discussion of statistics, is in terms of purpose and application rather than derivation or computation. Although the distinction between internal and external criteria is clearly made, the subsequent discussion of item analysis may be misinterpreted by the unsophisticated reader as implying that the former may be substituted for the latter. The intimate interdependence of test questions, directions, methods of administration, and even order of items in defining the nature of the total test is hinted at, but not made explicit enough to impress properly the beginning student of the field.

Chapter V presents a detailed discussion of certain types of performance tests for the measurement of job proficiency. Since this chapter was written by different authors, at a different time and originally for a somewhat different purpose than were the other chapters, it is incompletely integrated with the remainder of the book. The distinction between speed and power tests, although appropriate to performance tests as well as to others, has far wider application. The proper relation between reliability and validity, though briefly acknowledged, is not as clearly made here as it is in Chapter IV. The inexperienced reader may improperly infer that reliability is a desirable end in itself rather than a means to the real objective of valid measurement. Similarly, the discussion of internal weighting is not wholly consistent with, and is in general inferior to, that in Chapter IV. The suggestions for objectifying performance ratings and the detailed (sometimes too detailed) consideration of the step-by-step process of constructing a performance test will prove helpful to many.

Since the book has been written for practical use, concessions are occasionally, though rarely, made to expediency. Procedures are given for situations in which something must be done, but under such limiting conditions that nothing scientific can be done. There is always the difficult question, when the patient needs an appendectomy but stipulates that the incision must be in the hip so that no scar will show through a bare-mudriff sun suit, whether the physician should prescribe an ice pack.

One last point remains. The book may give the impression that the thoroughly sound techniques described characterize the bulk of examining by the U. S. Civil Service Commission. Within the re-

viewer's experience, written and performance tests are used for only a small proportion of the positions for which the Commission examines, and the techniques of construction and analysis described in this training manual are as yet applied to only a small part of these. It is hoped that publication of this book and its wide application will improve greatly the quality of examining in the federal service, in other public personnel agencies, and in the broader field of educational and psychological examining.

*Charles J. Monier*  
*Personnel Research Section*  
*Adjutant General's Office*



## MEASUREMENT ABSTRACTS<sup>1</sup>

Altus, William D. and Bell, Hugh M. "The Validity of a General Information Test for Certain Groups of Army Illiterates." *Journal of Consulting Psychology*, XI (1947), 120-132.

To determine more accurately whether the general capacity of previously screened Army inductees was adequate for military purposes despite their illiteracy, and to check on possible malingerers, the writers devised a 50-item test (called the *AB Information Test*) to supplement such officially sanctioned performance tests as the *Army Information Sheet*, the *Visual Classification Test*, the *Wells' Concrete Directions* or the *Block Counting Test*. Standardized on 1000 from the normal flow of men through an induction center, this test was found to have sufficient reliability and validity for the total range of intellect represented there. An opportunity for further validation and comparative studies on the *AB Information Test* were afforded the writers at an Army center for military and literacy training for classified illiterates. These inductees were composed of varied racial groups, including native-born Whites, Negroes, Indians, Filipinos, Chinese and bilingual as well as non-English-speaking Mexicans. *Vernon S. Tracht*.

---

Bennett, G. K. and Wesman, A. G. "Industrial Test Norms for a Southern Plant Population." *Journal of Applied Psychology*, XXXI (1947), 241-246.

From a population of adult white applicants for employment in a large southern industry, norms for the following tests are presented: the *Bennett Mechanical Comprehension Test*, Form AA; the *Revised Army Beta Examination*; the *Hand-Tool Dexterity Test*; and the *Minnesota Vocational Test for Clerical Workers*. Separate groupings have been made according to sex and whether the applications were for office or plant work. *LeRoy S. Burwen*.

---

Cattell, Raymond B. "Confirmation and Clarification of Primary Personality Factors." *Psychometrika*, XII (1947), 197-220.

In connection with a study bridging rating, questionnaire, and objective test factors, confirmation was sought with respect to the twelve personality factors previously found for young adult men. Variables were chosen to clarify and to discriminate the nature of related factors. Ratings of and by 373 students were obtained, and the present study describes the separate factorization for the 133

<sup>1</sup> Edited by Forrest A. Kingsbury.

men among them. Factorization yielded eleven factors, of which, on "blind" rotation for simple structure, 9 or 10 proved to be identical with those of the previous study. A new factor *M* is described. (Courtesy *Psychometrika*.)

Cawley, Sister Anne Mary. "A Study of the Vocational Interest Trends of Secondary School and College Women." *Genetic Psychology Monographs*, XXXV (1947), 187-247.

The *Strong Vocational Interest Blank for Women*--Form W B for Students--and the *Bernreuter Personal Inventory* were administered to 203 women students classified either as college juniors, college freshmen, or secondary-school juniors. Two years later a retest was given to 145 of the same subjects, all but one of the first group being at this time vocationally located, the majority of the second group still students, and half of the third group still students. Data analyzed include relationships between intelligence and measured interest tests and between measured interests and measured personality traits, ratings of stability and permanence of measured interests, relationship between vocational interest scores and achievement test scores or choice of college majors, relationship of vocational placement to personal satisfaction and to expressed and measured interests, and other measures. *Frances Smith*.

Dorkey, Margaret and Amen, Elizabeth W. "A Continuation Study of Anxiety Reactions in Young Children by Means of a Projective Technique." *Genetic Psychology Monographs*, XXXV (1947), 139-183.

A series of 14 pictures, differentiated for boys and girls with regard to the sex of the central figure, were presented to 24 children ranging in age from three years and ten months, to five years and eight months, with an IQ spread of from 75 to 144. The picture-scenes covered familiar areas of experience, seven of the pictures being designated emotionally "equivocal." An analysis of the responses indicates in the children studied a frequent association of anxiety with specific areas of experience, especially with the field of child-child relationships, the "equivocal" pictures discriminating to a high degree between children with high and with low anxiety ratings. Test reliability was .95 for split-half correlation, and .22 for retest correlation, the latter figure being interpreted as showing sensitivity of the test to psychological changes in the child. *Frances Smith*.

Duncan, Acheson J. "Some Comments on the Army General Classification Test." *Journal of Applied Psychology*, XXXI (1947), 143-149.

The author states that the major emphasis of the AGCT is on spatial thinking and quantitative reasoning. The speed factor in the test is of some importance also. Probably those who like to gamble will have some advantage. Skipping a question on the test paper by

mistake is heavily penalized. The test is to some extent successful in distinguishing between high learning ability and very high learning ability. A fair degree of correlation is found between AGCT scores and amounts of formal education. In one group of 100 men, an average increase in standard score in a retest on the AGCT was 11.25 points. The correlation between AGCT score and rank in the service is low, although correlation between rank and length of time in the service is fairly high. *LeRoy S. Burton.*

---

Festinger, Leon. "The Treatment of Qualitative Data by 'Scale Analysis.'" *Psychological Bulletin*, XLIV (1947), 149-161.

The author reviews the published literature and some unpublished literature on the technique of "scale analysis" as developed by Guttman and others. Specific reference is made to its applicability to determining the existence of uni-dimensionality, and a general evaluation of the technique is given. *LeRoy S. Burton.*

---

Fowler, H. M. "The Consistency of the Items of an Activity Preference Blank." *Psychometrika*, XII (1947), 221-232.

The results of an experiment to obtain data on the consistency of the items of two forms of an *Activity Preference Blank* are presented. Both Form I and Form II, which was a revised edition of Form I, were administered twice, so consistency data are available for both forms. A sub-item is said to be consistent if a high proportion of men marked it the same way, *M* for preferred "Most" and *L* for preferred "Least," on both administrations of the test. The data of the experiment were investigated to see what happens to the consistency of sub-items when the items are changed in context, when the number of sub-items in an item is reduced, and when the time-interval between the administration and the re-administration of the test is increased. The author also gives data on the consistency of the responses made to particular combinations of sub-items and data on item consistency when all sub-item combinations are taken into consideration. (*Courtesy Psychometrika.*)

---

Guetzkow, Harold and Brozek, Josef. "Intellective Tests for Longitudinal Experiments on Adults." *American Journal of Psychology*, LX (1947), 350-366.

In selection of intellective tests suitable for periodic assessment of intellective performance capacity, the Laboratory of Physiological Hygiene at the University of Minnesota assembled a battery of six tests factorially distinct, and suitable for being practiced to a plateau. Equivalent test forms were provided to avoid memorization of responses. When administered to trained subjects undergoing nutritional stresses, the battery in general indicated only a slight disturbance of intellective functions. The more marked impairment occurring on a limited number of tests offers evidence in support of the hypothesis of relatively independent "mental factors." *Frances Smith.*

Guttman, Louis. "On Festinger's Evaluation of Scale Analysis." *Psychological Bulletin*, XLIV (1947), 451-465.

The author presents further information on the technique of "scale analysis" and refutes and clarifies certain points raised by Festinger in a previous article. He points out that Festinger's article is admittedly tentative and incomplete since it is not based upon all of the information available. *Leroy S. Burwen.*

Jensen, M. B. and Rotter, J. B. "The Value of Thirteen Psychological Tests in Officer Candidate Screening." *Journal of Applied Psychology*, XXXI (1947), 312-322.

The following tests were administered to 1492 students in an Officer Candidate School: (1) *Personnel Test*, (2) *Stanford Achievement Test*, (3) *Test of Mechanical Comprehension*, Form AA, (4) *Army General Classification Test*, (5) *Speed of Substitution Test*, (6) *Rhythm Test*, (7) *P-Inventory*, (8) *C-Inventory*, (9) *Level of Aspiration Test*, (10) *Thematic Apperception Test*, (11) *Honesty Tests* (Circles and Squares), (12) *Harrower-Erickson Multiple Choice Rorschach Test*, and (13) *Vocational Interest Schedule*. Results indicated that the most efficient combination of tests included Arithmetic Computation, combined Paragraph Meaning and Word Meaning from the new *Stanford Achievement Test*, and the *Wonderlic Personnel Test*. *Leroy S. Burwen.*

Kendall, W. E. "The Occupational Level Scale of the Strong Vocational Interest Blanks for Men." *Journal of Applied Psychology*, XXXI (1947), 283-287.

From a class of male university freshmen to whom the Strong blank had been administered, three groups of 100 men each were chosen on the basis of OI scores. Median raw scores for the three groups were 117, 18, and -55. The academic ability of the subjects was measured by the *Ohio State University Psychological Test*, Form 21, and their scholastic achievement by the first semester hour-point ratio. The three groups were found to differ significantly with respect both to achievement and to ability. When the effect of academic ability was partialled out, the groups were found to differ with respect to achievement. The role of motivational factors in scholastic success is suggested as a problem for further study. *Frances Smith.*

Lawshe, C. H., Jr. and Forster, Max H. "Studies in Projective Techniques: I. The Reliability of a Multiple Choice Group Rorschach Test." *Journal of Applied Psychology*, XXXI (1947), 199-211.

A modified form of the *Harrower-Erickson Amplified Multiple Choice Test* was administered to 655 college students, 371 men and 284 women. Split-half reliability was .61 for first choice scores and .40 for alternate choice scoring. The results of further analysis indicate the unreliability of the standard scoring and the need for

separate male and female scoring keys. A classification of some of the items is questionable. The intelligence factor has little effect on the variability of scores. It is also thought that an increase in the number of ink-blots might render the test more useful in individual applications. *Leroy S. Burton.*

---

Lawshe, C. H., Jr. and Mayer, J. S. "Studies in Item Analysis: 1. The Effects of Two Methods of Item Validation on Test Reliability." *Journal of Applied Psychology*, XXXI (1947), 271-277.

Two methods of item validation, the Planagan "r" method and the D-value method based on Lawshe's nomograph, were subjected to statistical comparison in order to discover which yielded items that created a more reliable test, as measured by the split-half method. The results indicate that the "r" method yields approximately the same reliability coefficients for the different length tests. At the same time the D-value method gives fluctuating reliabilities, but yields increasingly higher reliabilities for tests of greater moderate length. *Leroy S. Burton.*

---

Lewis, John A. "Kuder Preference Record and MMPI Scores for Two Occupational Groups." *Journal of Consulting Psychology*, XI (1947), 194-201.

Two occupational groups consisting of 50 white male insurance agents and 50 white female social workers were given the *Kuder Preference Record* and the *Minnesota Multiphasic Personality Inventory*, the purpose being to investigate the possibility of a relationship between measured occupational interests and personality tendencies, and to discover something about the extent of this relationship if it should exist. Results indicate that such a relationship is present, and it appears to be inversely proportional when the occupation in which the person is engaged is considered, i.e., persons relatively uninterested in the type of work they are doing tend to make more abnormal scores on the MMPI than those relatively interested. The desirability of further research is indicated, according to the writer, with various other occupational groups in which ability, job satisfaction, and the degree of vocational adjustment are made simultaneously. *Vernon S. Tracht.*

---

MacPhee, H. M., Wright, H. F., and Cummings, S. B., Jr. "The Performance of Mentally Subnormal Rural Southern Negroes on the Verbal Scale of the Bellevue Intelligence Examination." *Journal of Social Psychology*, XXV (1947), 217-229.

Four hundred and thirty-two rural Southern Negroes with a mean age of 21 years and a mean educational level of three grades, all referred to a naval neuropsychiatric unit because of suspected mental retardation, were given the *Verbal Scale of the Bellevue Intelligence Examination*. An analysis of weighted scores provided information

as to "patterning" of subtest scores, changes in subtest performance at successive levels of ability, and subtest intercorrelations. It is concluded from the results that the Verbal Scale is a dependable measure of the general intellectual ability of the group studied. *Frances Smith.*

---

Marks, Eli S. "Sampling in the Revision of the Stanford-Binet Scale." *Psychological Bulletin*, XLIV (1947), 413-429.

The author presents an analysis of the effects of cluster sampling in psychological research, stating that it almost always results in an increase in sampling error as compared with unrestricted random sampling of the same number of cases. He points out that the intra-class correlation of the population cannot be ignored when "cluster" methods of sampling are used. *Leroy S. Barwen.*

---

Medland, Francis F. "An Empirical Comparison of Methods of Communalities Estimation." *Psychometrika*, XII (1947), 101-109.

This paper shows the formulation of nine methods of estimating the unknown communalities. Each of these methods has been used on experimental data and the results have been tabulated for comparison. They show that the most accurate approximations are obtained from the *Centroid No. 1* and the *Graphical* methods. (Courtesy *Psychometrika*.)

---

Pennington, L. A. "The Serial Sevens Test as a Psychometric Instrument." *American Journal of Orthopsychiatry*, XVII (1947), 488-499.

The *Verbal Battery of the Wechsler-Bellevue Scale* and the *Serial Sevens Test* were administered to 201 White and 77 Negro adult males with IQ's ranging from 47 to 94. The correlation of Bellevue verbal IQ scores with Sevens Test scores was found to be too low for predictive value, as were correlations of Sevens Test scores with each of the Bellevue subtests. It was further found that equal success on the Sevens Test was attained by individuals with an IQ range of from 55 to 94, and that no pattern of sequence in this test is typical or diagnostic of the ament. The *Serial Sevens Test* is concluded to be restricted in its range of applicability and its wholesale use in screening examinations is ill-advised. *Frances Smith.*

---

Prados, Miguel and Fried, Erdita G. "Personality Structure of the Older Age Groups." *Journal of Clinical Psychology*, III (1947), 113-120.

The Rorschach technique was used in the personality study of a group of 35 men and women aged 50 to 80, who were subdivided into age groups of 50 to 60, 61 to 70, and 71 to 80. An analysis of results in the various scoring categories suggests impoverishment of the creative intellectual faculties with increasing age, accompanied by lessen-

ing anxiety in response to awareness of inadequacy. Capacity for emotional responsiveness to the environment appears to diminish, with a corresponding tendency to re-occurrence of primitive infantile emotional manifestations. *Frances Smith*

---

Rabin, Albert I. "Vocabulary and Efficiency Levels as Functions of Age in the Babcock Method." *Journal of Consulting Psychology*, XI (1947), 207-211.

*Babcock Test* records of 404 state hospital patients subdivided diagnostically into seven major classes and representing a seven-decade age range, were analyzed to ascertain the effect of age on the efficiency index. All of the diagnostic groups showed lower efficiency ratings at the higher age levels, accompanied by a tendency of vocabulary to rise with age. It is concluded that the establishment of age norms is important in increasing the usefulness and validity of the *Babcock Test*. *Frances Smith*

---

Sarason, Seymour B. and Sarason, Esther Kroop. "The Discriminatory Value of a Test Pattern with Cerebral Palsied, Defective Children." *Journal of Clinical Psychology*, III (1947), 141-147.

Eighteen cerebral palsied subjects, with average C.A.'s and M.A.'s of 19.5 and 8.1 years, respectively, were given the 1937 Binet (Form L), the *Arthur Performance Scale*, the Rorschach, and an EEG examination to study the diagnostic value of the obtained test pattern and to compare these results with a previous study on familial defectives. The results indicated that the test pattern aids in discriminating between those whose cerebral damage is cortical and the cause of the mental deficiency, and those with sub-cortical damage which is probably not the cause of the deficiency. The EEG, being a sensitive test for cortical disfunctioning, establishes the diagnostic value of the test pattern by the substantial agreement found between the two. As was true of the familials, the cerebral palsied apparently are not a homogeneous group from the standpoint either of cause or of mental functioning. *Vernon S. Tracht*

---

Super, Donald E. "The Kuder Preference Record in Vocational Diagnosis." *Journal of Consulting Psychology*, XI (1947), 184-193.

Now that the Kuder has apparently "come of age," it having been eight years since its publication, this article reviews the accumulated data regarding its value in vocational counseling and selection. A comparison of the Kuder with the *Strong Vocational Interest Blank* is made throughout, not only in the brief description of its development and standardization but in the discussion of its validity as well. While the two inventories are related in some expected ways and not in others, the writer suggests that, in general, the differences between them are significant for diagnostic purposes: the Kuder measures a subject's relative interest in a type of activity, whereas

the Strong measures the similarity of a subject's varied interests to those of persons who have been successful in an occupation. *Vernon S. Tracht*.

#### ADDITIONAL ARTICLES NOT ABSTRACTED

- Alexander, Howard W. "The Estimation of Reliability When Several Trials Are Available." *Psychometrika*, XII (1947), 79-99.
- Ash, P. "The Discrepancy Between Reported Schooling and Testing Scholastic Ability Among Adolescent Delinquents." *Journal of Applied Psychology*, XXXI (1947), 323-328.
- Borow, Henry. "Current Problems in the Prediction of College Performance." *Journal of the American Association of Collegiate Registrars*, XXII (Oct. 1946), 14-26.
- Brower, Daniel. "The Relation Between Intelligence and Minnesota Multiphasic Personality Inventory Scores." *Journal of Social Psychology*, XXV (1947), 243-245.
- Burt, Cyril. "Symposium on the Selection of Pupils for Different Types of Secondary Schools. I. A General Survey." *British Journal of Educational Psychology*, XVII (1947), 57-71.
- Cofer, Charles N. "Psychological Test Performance Under Hyocine: A Case of Post-Infectious Encephalopathy." *Journal of General Psychology*, XXXVI (1947), 221-228.
- Cross, Orrin H. "Braille Edition of the Minnesota Multiphasic Personality Inventory for Use with the Blind." *Journal of Applied Psychology*, XXXI (1947), 189-198.
- Damrin, Dora E. "A Study of the Truthfulness with Which High School Girls Answer Personality Tests of the Questionnaire Type." *Journal of Educational Psychology*, XXXVIII (1947), 223-231.
- Doll, Edgar A. "Note on the Age Placement of Year-Scale Tests." *Journal of Consulting Psychology*, XI (1947), 144-147.
- Dunkel, Harold B. "The Effect of Personality on Language Achievement." *Journal of Educational Psychology*, XXXVIII (1947), 177-182.
- Fischelli, V. R. and Welch, L. "The Ability of College Art Majors to Recombine Ideas in Creative Thinking." *Journal of Applied Psychology*, XXXI (1947), 278-282.
- Fox, V. "A Study of the Promotion of Enlisted Men in the Army." *Journal of Applied Psychology*, XXXI (1947), 298-305.
- Garfield, Sol L. "The Shipley-Hartford Retreat Scale as a Quick Measure of Mental Status." *Journal of Consulting Psychology*, XI (1947), 148-150.
- Garfield, Sol L. "A Note on Patterson's Article, 'The Wechsler-Bellevue Scale as an Aid in Psychiatric Diagnosis'." *Journal of Clinical Psychology*, III (1947), 198-200.
- Gerberich, John B. "A Study of the Consistency of Informant Responses to Questions in a Questionnaire." *Journal of Educational Psychology*, XXXVIII (1947), 299-306.



- Gulde, Carl J. and Roy, Howard L. "A Note on the Scoring of the Minnesota Multiphasic Personality Inventory." *Journal of Consulting Psychology*, XI (1947), 221-222.
- Gurvitz, Milton S. "The Intelligence Factor in Psychopathic Personality." *Journal of Clinical Psychology*, III (1947), 194-196.
- Johnson, Palmer O. and Hunt, Cecil. "On Determining Three Dimensional Regions of Satisfaction." *Journal of Experimental Education*, XV (1947), 203-212.
- Klebanoff, Seymour G. "Personality Factors in Symptomatic Chronic Alcoholism as Indicated by the Thematic Apperception Test." *Journal of Consulting Psychology*, XI (1947), 111-119.
- Kobler, Frank J. "The Measurement of Improvement Among Neuro-Psychiatric Patients in an Army Convalescent Facility." *Journal of Clinical Psychology*, III (1947), 121-128.
- Leverett, Hollis M. "Notes on the Use of the Normal Distribution in Psychometrics." *Journal of Educational Psychology*, XXXVIII (1947), 283-289.
- Nardi, Noah. "A Test to Measure Aptitude in the Hebrew Language." *Journal of Educational Psychology*, XXXVIII (1947), 167-176.
- Oxlade, M. N. and Walker, K. F. "A Note on Adapting the Minnesota Rate of Manipulation Test to Factory Use." *Journal of Applied Psychology*, XXXI (1947), 247-248.
- Rashkis, Harold A. "The Psychometric Analysis of a Diagnostic Problem." *American Journal of Orthopsychiatry*, XVII (1947), 529-532.
- Sarason, Seymour B. and Potter, Elmer H. "Color in the Rorschach and Kohs Block Designs." *Journal of Consulting Psychology*, XI (1947), 202-206.
- Seashore, H. G. "The Improvement of Performance on the Minnesota Rate of Manipulation Test When Bonuses Are Given." *Journal of Applied Psychology*, XXXI (1937), 254-259.
- Seashore, H. G. "The Superiority of College Students on the Minnesota Rate of Manipulation Test." *Journal of Applied Psychology*, XXXI (1947), 249-253.
- Stright, I. L. "Some Factors Affecting College Success." *Journal of Educational Psychology*, XXXVIII (1947), 232-240.
- Swineford, Frances. "Growth in the General and Verbal Bi-Factors from Grade VII to Grade IX." *Journal of Educational Psychology*, XXXVIII (1947), 257-272.



## THE CONTRIBUTORS

Philip H. DuBois—Ph.D., Columbia University, 1932. Assistant Professor of Psychology, University of Idaho, 1933–1935. Member of the Faculty of the University of New Mexico, 1935–1946. Aviation Psychologist, San Antonio Aviation Cadet Center, Texas; Jefferson Barracks, Missouri; the North African Theater of Operations; Headquarters, AAF Training Command, Fort Worth, Texas, 1942–1946. Professor of Psychology and Director of the University Testing Program, Washington University, 1946–.

Robert L. Ebel—Ph.D., University of Iowa, 1947. High-school teacher in Iowa, 1932–1938. High-school teacher and Principal, The Edison Institute, Dearborn, Michigan, 1938–1944. Assistant in Educational Measurement, 1944–1945; Assistant Director, University Examinations Service, 1945–; Assistant Professor of Education, 1947–, University of Iowa. Author of articles on the teaching of science. Co-author of several educational tests.

Vernon Lemmon—Ph.D., Columbia University, 1928. Research Associate, University of California Institute of Child Welfare, 1928. Associate Professor of Psychology, Washington University, 1928–. Acting Associate Professor of Physics, Washington University, 1942–1944. Associate Member, American Psychological Association. Member, American Physical Society, Tau Beta Pi, Sigma Xi.

Louis Long—Ph.D., Columbia University, 1937. Assistant, Department of Psychology, Columbia University, 1936–1937. Research Assistant and Instructor, Department of Pediatrics, 1937–1941; Instructor, Department of Psychology, 1940–1941; Columbia University. Instructor, Department of Psychology, Hunter College, 1940–1941. Instructor, Student Personnel Bureau, 1941–1945; Assistant Professor, 1946–; Director, Division of Testing and Guidance of Department of Student Life, 1946–; Director of Vocational Advisement Unit, 1946–, City College of New York. Author of articles in *Journal of Consulting Psychology*, *Journal of Applied Psychology*, *Journal of Experimental Education* and other journals. Member, American Psychological Association, National Vocational Guidance Association, American Association for the Advancement of Science, Sigma Xi.

Harold B. Pepinsky—Ph.D., University of Minnesota, 1946. Teacher and Research Assistant, 1938–1942; Instructor, College of

Education, summer 1943; Administrative Fellow, Dean of Students' Office, 1942-1944, University of Minnesota. Assistant Director, Guidance Bureau, and Assistant Professor of Education, University of Kansas, 1944-1945. Assistant Professor, Institute of Counseling, Testing and Guidance; Assistant Professor, Department of Effective Living, Michigan State College, 1946-. Visiting Assistant Professor, New York University, 1947. Author of articles on counseling and co-author of *Financial Assistance for College Students*. Co-Editor of *Basic Values in Human Relationships*. Member, American College Personnel Association, National Vocational Guidance Association, Phi Delta Kappa, Society for the Psychological Study of Social Issues.

James D. Perry—M.A., Teachers' College, Columbia University, 1932. High-school principal, Glenpool, Oklahoma, 1922-1934. Statistician, Department of Student Life, City College of New York, 1935-. Author of articles in *School and Society*, *Journal of Psychology*, and *Journal of Educational Research*.

Harold G. Seashore—Ph.D., University of Iowa, 1933. Research Assistant, University of Iowa, 1933-1935. Lecturer in Psychology, University of Iowa, two summers. Professor of Psychology and Director of the Freshman Year, Springfield College, Mass., 1935-1942. Senior Research Associate, Project N-107, OSRD, 1942-1944. Associate Director and Director, Test Division, The Psychological Corporation, N. Y. City, 1944-. Author of articles on measurement, guidance, psychology of music, development of motor skills and learning. Author of several psychological tests. Fellow, American Psychological Association. Member, N. Y. State Association for Applied Psychology, Eastern Psychological Association, National Vocational Guidance Association, Sigma Xi, and Committee on Counseling of the National Council of the Y.M.C.A.

David Segel—Ph.D., Stanford University, 1931. Assistant Director of Research, Long Beach (Calif.) City Schools, 1924-1931. Educational Consultant for Tests and Measurements, U. S. Office of Education, 1931-. Consultant to the Office of the Secretary of War, 1944. Consultant to Military Government in Germany, April-May, 1947. Contributing Editor of the *Journal of Experimental Education*. Author of articles and bulletins on the construction and use of measurements for guidance purposes. Member, American Psychological Association, American Educational Research Association.

Naomi Stewart—Ph.D., Purdue University, 1944. Personnel Technician, Personnel Research Analyst, Personnel Research Section, AGO, 1944-1945 and 1946-1947. Instructor in Psychology, Brooklyn College (Eve. sessions), 1943-1945 and 1946-1947. Vocational Appraiser, Veterans Administration Center, Rutgers University, 1945-1946. Head of Research Department, Cooperative Test Service of the American Council on Education, 1947-. Lecturer in Psychology, Graduate School of Arts and Science, New York Uni-

versity, 1947-. Associate Member, American Psychological Association. Member, Psychometric Society, American Statistical Association.

**Dewey B. Stuit**--Ph.D., University of Illinois, 1934. Instructor in Psychology and Education, Carleton College, 1934-1936. Associate Professor of Educational Psychology and Measurements, University of Nebraska, 1936-1938. Associate Professor of Psychology and Director, Personnel Research, University of Iowa, 1938-1943. On military leave, U. S. Navy, 1943-1946. Professor of Psychology and Director of Student Counseling Office, University of Iowa, 1946, and Coordinator of Student Personnel Services, 1947-. Director of Studies, Association of American Medical Colleges, 1947-. Author of articles on tests and student personnel work. Editor, *Personnel Research and Test Development in the Bureau of Naval Personnel*, Princeton Press, 1947. Fellow, American Psychological Association. Member, American College Personnel Association, Midwestern Psychological Association, Phi Delta Kappa, Phi Beta Kappa, Sigma Xi.

**Donald E. Super**--Ph.D., Columbia University, 1940. Assistant Employment Secretary, Y.M.C.A., Cleveland, Ohio, 1932-1935. Director, Cleveland Guidance Service, Cleveland, Ohio, 1935-1936. Research Assistant, Teachers College, Columbia University, 1936-1938. Associate Professor of Educational Psychology and Director, Student Personnel Bureau, Clark University, 1938-1942. Aviation Psychologist, Army Air Forces, 1942-1945. Associate Professor of Education, Department of Guidance, Teachers College, Columbia University, 1945-. Author of articles in psychological and educational journals and of aviation psychology research reports. Author of two books on measurement and guidance. Fellow, American Psychological Association (President, Division of Consulting Psychology). Member, New York State Association of Applied Psychologists, Eastern Psychological Association, National Vocational Guidance Association, and American College Personnel Association.

**Claude Edward Thompson**--Ph.D., Ohio State University, 1939. Assistant Personnel Director, Ohio Fuel Gas Co., Columbus, Ohio, 1936-1938. Instructor and Research Associate, Ohio State University, 1936-1939. Psychologist, U. S. Public Health Service, U.S.P.H.S. Hospital, Lexington, Kentucky, 1939-1940. Assistant Professor, Department of Psychology, Northwestern University, 1940-1945. Training Director, Montgomery Ward, Chicago, 1943. Director of Aptitude Research, U. S. Naval Training School (Radio), Technological Institute, Northwestern University, 1942-1943. Associate Professor and Associate Director, Institute of Counseling, Testing and Guidance, Michigan State College, 1945-1946. Professor of Business and Industrial Psychology and Director of Adult Testing and Guidance, University of Omaha, Nebraska, 1946-.

Kenneth W. Vaughn—Ph.D., University of Iowa, 1941. Research Associates, University of Iowa, 1941-1942. Project Director, 1942-1943; Measurement and Guidance Project in Engineering Education, 1943-. Graduate Record Office of the Carnegie Foundation for the Advancement of Teaching, Associate Director, 1945-1946; Director, 1946-. Associate Director of the Cooperative Test Service of the American Council on Education, 1945-. Author of articles on measurement and research problems, engineering education, and general education. Member, Association for the Advancement of Science, American Society for Engineering Education, Institute of Mathematical Statistics, National Educational Research Association, and Phi Delta Kappa.

D. A. Worcester—Ph.D., Ohio State University, 1926. Supervisor of Elementary Schools, Zamboanga, P. I., 1913-1914. Principal, Cheyenne High School, Cheyenne, Wyoming, 1914-1918. Professor of Psychology, University of New Mexico, 1918-1920. Professor of Psychology and Director of Bureau of Educational Tests and Measurements, Kansas State Teachers College, 1920-1924. Instructor in Psychology, Ohio State University, 1924-1926. Associate Professor of Psychology, Ohio University, 1926-1927. Associate Professor of Educational Psychology, 1927-1928; Professor of Psychology, 1928-; Chairman of Dept. of Psychology, 1934-, University of Nebraska. American Army University, Biarritz, France, 1945-1946. Co-author of two books and author of various articles in educational and psychological journals. Fellow, American Psychological Association, American Educational Research Association, American Association for the Advancement of Science, Phi Beta Kappa, Sigma Xi, Phi Delta Kappa.

Gilbert C. Wrenn—Ph.D., Stanford University, 1932. Vocational Counselor, Stanford University, 1928-1936. Associate Director, General College, and Associate Professor of Educational Psychology, 1936-1938. Professor of Educational Psychology, University of Minnesota, 1938-. On military leave, 1942-1946, to serve in the Bureau of Naval Personnel and Pacific area as Personnel Officer in the U. S. Army. Associate, American Youth Commission, 1939-1941; Consultant, Student Personnel Teacher Education Commission of the American Council on Education, 1939-1942. President, National Vocational Guidance Association; Vice-President, American College Personnel Association; Vice-President, Council of Guidance Personnel Association, 1946-. Author and co-author of *Student Personnel Problems*, *Studying Effectively*, *Aids to Group Guidance*, *Time on Our Hands*, and of numerous articles.

